Summer 8-15-2019

Multi-omic Understanding of the Evolution of Xenobiotic Tolerance in Bacterial Isolates and Communities

Tayte Paul Campbell
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Plant and Microbial Biosciences

Dissertation Examination Committee:
Gautam Dantas, Chair
Arpita Bose
Andrew Kau
Audrey Odom-John
Himadri Pakrasi
Fuzhong Zhang

Multi-omic Understanding of the Evolution of Xenobiotic Tolerance in Bacterial Isolates and Communities

by
Tayte P. Campbell

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Gautam Dantas, for his mentorship and guidance. Gautam has created a wonderful lab culture of rigorous science, friendship, and collaboration that has made for an unparalleled graduate school experience.

I would like to thank my lab members for their feedback and camaraderie throughout my graduate school education. They always made me look forward to working in lab and provided help and assistance when experiments weren't working out.

A sincere thank you to my thesis committee members: Arpita Bose, Andrew Kau, Audrey Odom-John, Himadri Pakrasi, and Fuzhong Zhang. Thank you for taking time out of your busy schedules to meet with me and provide insightful advice and feedback.

I thank my family for being there for me since day one and for stressing the importance of education. I am grateful for the examples you set for me. Thank you for nurturing my curiosity and love of learning.

I cannot thank my wife enough for her support. Tenery Campbell has given me unwavering encouragement as I follow my dreams. I'm grateful for her enduring love and for all that she has sacrificed for me. I'm grateful for my daughter, Sophia Campbell, for the happiness and unconditional love she has brought into my life.

Tayte P. Campbell

Washington University in St. Louis

August 2019

Dedicated to my wife and family.

ABSTRACT OF THE DISSERTATION

Multi-omic Understanding of the Evolution of Xenobiotic Tolerance in Bacterial Isolates and Communities

by

Tayte P. Campbell

Doctor of Philosophy in Biology and Biomedical Sciences

Plant and Microbial Biosciences

Washington University in St. Louis, 2019

Professor Gautam Dantas, Chair

Xenobiotic compounds are any chemicals that are released into an environment by human action and that occur at concentrations higher than found naturally. Xenobiotics, including aromatic compounds and antibiotics, are recalcitrant to degradation because they are often toxic or mutagenic. Despite this toxicity, bacteria account for a large portion of xenobiotic degradation in the environment. Bacteria are able to adapt to these foreign chemicals, gaining increased levels of tolerance and increased rates of xenobiotic degradation. On the strain level, increased tolerance can be caused by mutations in individual cells or through the acquisition of genes from other cells. At the community level, xenobiotics select for naturally resistant bacterial often resulting in an increase in genes involved with xenobiotic tolerance. The goal of my thesis was to (1) understand how bacterial strains evolve increased tolerance to toxic aromatics xenobiotics through adapted evolution and (2) how microbial communities in mammalian guts are altered due to xenobiotic selection pressure.

To determine the microbial adaptations that occur in bacterial strains exposed to xenobiotics, I studied the changes that occurred in the genome and transcriptome of *Rhodococcus*

opacus PD630 when grown over several generations on increasing concentrations of toxic aromatic compounds. Chemical pretreatment of lignocellulose as a first step in biofuel production results in the creation of monomeric sugars and toxic xenobiotic compounds. Bacterial conversion of the resulting moiety to biofuel precursors is one of the most cost-effective methods of biofuel production. However, the toxic aromatic compounds created from lignocellulose pretreatment inhibit bacterial growth and reduce overall productivity. *R. opacus* is a bacterial strain that naturally has a high tolerance to aromatic compounds and optimization of this bacteria can improve process efficiency. By analyzing 35 *R. opacus* strains adapted on 6 different compounds or compound mixtures, I show that adapted strains demonstrated up to 1900% improvement in final cell densities. I found no mutations in xenobiotic degradations genes for the adapted strains, but I found several mutations in genes that are involved with oxidation-reduction reactions that may assist in degradation. These results demonstrate that bacterial strains can gain increased xenobiotic tolerance by fine-tuning metabolic pathways indirectly related to xenobiotic degradation.

To determine the effects of xenobiotic selection on microbial communities, I studied the gut microbiome of humans, captive chimpanzees, and captive gorillas that have received antibiotic treatment compared to the microbiome of wild chimpanzees and gorillas that have never received antibiotic treatment. I found that antibiotic treatment was correlated with higher richness and abundance of antibiotic resistance genes. In addition, the microbiome and resistome in captive apes were more similar to that of humans than to the wild apes, despite differences in host species and large geographic distances between the human and captive apes. Together, these results suggest that host lifestyle, including diet and antibiotic treatment, is more influential in microbiome composition than host species and geographic proximity. I also identified a number of novel antibiotic resistance genes for which further investigation is warranted.

Chapter 1

Introduction

1.1 Xenobiotics

Xenobiotics have been produced in enormous quantities over the past century as a result of industrial activity [1]. In a broad sense, xenobiotics are any compounds that are foreign to the natural environment or found at much higher concentrations than usual [2]. They may either be artificially synthesized (e.g. chlorinated biphenyls, alkylbenzyl sulphonates, and halocarbons), natural products used at high concentrations (e.g. most antibiotics), or created by incomplete combustion of organic materials (e.g. phenolics and polycyclic aromatic hydrocarbons (PAHs)) [3-6]. Xenobiotic compounds are often toxic and recalcitrant to biodegradation, leading to contamination of the environment and bioaccumulation in living organisms [7]. Bacterial degradation of xenobiotics can be harnessed to either degrade unwanted pollutants to harmless byproducts through bioremediation, or to convert abundant xenobiotics to value added chemicals [8, 9]. In the case of antibiotics, targeted xenobiotic toxicity is utilized to kill pathogenic microbes in host organisms without harming the host [10]. However, microorganisms are becoming more resistant to antibiotics and increasingly able to metabolize antibiotic compounds to harmless byproducts [11, 12]. Understanding the microbial processes necessary for xenobiotic tolerance and degradation will allow for the engineering of improved bacterial strains for bioremediation and

industrial production, while also providing key insight to combat the spread of antibiotic resistance.

1.2 Mechanisms of Xenobiotic Toxicity

Many aromatic xenobiotics cause toxicity through the formation of quinones in the cell [13]. Quinones, fully conjugated cyclic compounds, are toxic due to the creation of free radicals that react with cell components [13, 14]. For example, free radicals damage cells by causing DNA mutations and by oxidizing lipids, amino acids, and carbohydrates [15, 16]. Quinones are particularly toxic because when they interact with cellular enzymes it leads to a phenomenon termed futile redox cycling. Specifically, futile redox cycling occurs when a cytochrome-P450 reductase uses NADPH to reduce a quinone to semiquinone by adding one electron [17]. The resulting semiquinone rapidly reacts to produce a superoxide anion by passing the radical electron onto an oxygen molecule resulting in the formation of the original quinone, restarting the cycle [13]. The generated superoxide anion can then damage cell components with its highly reactive electron radical. Futile cycling thus leads to a rapid increase in harmful radical species within the cell while also depleting NADPH reservoirs within the cell [18]. Xenobiotics that cause toxicity through this mechanism are typically toxic to both prokaryotic and eukaryotic cells due to the non-specific nature of free radical reactions [19].

In contrast, antibiotic compounds are targeted to specific cell types and typically bind to targeted proteins or lipids [20]. This specific activity allows for their use to clear bacterial infections in humans and animals without causing harm to host cells and tissues [10, 21, 22]. Antibiotics target cellular components that are conserved across the bacterial domain of life and that are either absent or sufficiently different in host cells [23]. Current antibiotics only target a small number of cellular components and pathways such as bacterial ribosomes, lipid membranes,

the cell wall synthesis pathway, the folate synthesis pathway, and proteins involved in DNA replication and maintenance [24]. Antibiotics bind to specific proteins and lipids resulting in inhibition of the protein activity when bound to the former or, when bound to the latter, disruption of the cell membrane.

1.3 Xenobiotic Tolerance

Bacterial tolerance to xenobiotic compounds in individual isolates is typically achieved through one of four common mechanisms: 1) xenobiotic degradation or modification, 2) alteration or protection of the xenobiotic target, 3) active efflux of the xenobiotic, or 4) reduced permeability of the xenobiotic into the cell [10]. Since these resistance mechanisms cause resistance in different ways, the presence of more than one in a cell can result in higher levels of resistance [25]. For example, if a bacterial strain encodes both a xenobiotic degradation enzyme and has reduced permeability, then fewer xenobiotic compounds enter the cell and those that do are degraded [25]. When one mechanism becomes oversaturated, the other mechanism can compensate and keep xenobiotic concentrations low within the cell. In fact, multiple resistance mechanisms may be necessary for tolerance if both mechanisms have low specificity to the xenobiotic [25].

Bacterial enzymes have evolved over millennia to metabolize substrates that are commonly found in their natural environment [26]. Since xenobiotic compounds contain uncommon chemical structures or properties, bacteria often don't intrinsically have enzymes that are capable of their degradation [27]. However, upon exposure to these unnatural compounds, bacteria are able to adapt relatively quickly and increase their levels of tolerance and degradation [28]. In fact, communities pre-exposed to specific xenobiotics or structurally similar compounds are able to increase their degradation rates up to 1,000-fold higher than unexposed populations [29]. Bacteria are able to adapt by either acquiring necessary genes from other bacteria through a mechanism

termed horizontal gene transfer or through acquired mutations that alter enzymatic specificities or create novel metabolic activities [2, 26].

In complex microbial communities, xenobiotic exposure selects for bacterial species that are able to tolerate the xenobiotic [26, 30]. One typical result of this selection pressure is reduced richness and diversity in bacterial communities as non-tolerant strains are often eliminated [31]. Another complication is that of these tolerance mechanisms, only microbial degradation reduces the amount of xenobiotic in the environment. As a result, degradation benefits bacteria with other resistance mechanisms and even bacteria that would otherwise be susceptible to the xenobiotic. These bacteria are termed cheaters since they benefit from the activity of other organisms [32]. This can lead to interesting community dynamics when the fitness cost of maintaining resistance genes is higher than the fitness cost of antibiotic susceptibility [33]. Paradoxically, the end result of this is selection against the individual strains that are performing xenobiotic degradation. Due to this wide array of community responses, it's necessary to analyze the taxonomic changes along with the xenobiotic tolerance mechanisms in the community to begin to tease apart the community dynamics associated with xenobiotic selection pressure. DNA sequencing of microbial communities allows for both the identification of bacterial taxa and individual genes [34-36]. However, in complex microbial communities it can be difficult to associate specific genes with their host species, especially with the short DNA sequencing reads that are common today. With these limitations, correlational findings can be inferred statistically but may need longer sequencing lengths and co-culturing of simplified communities to further elucidate these dynamics.

1.4 Lignocellulose Degradation for Biofuel Production

Even though biological process may occur at very slow rates, microbial conversion of xenobiotics provides a cost effective solution for bioremediation or industrial conversion processes [37, 38]. For example, lignocellulosic biomass, the most abundant biological compound on the planet, is often targeted for biological conversion to biofuels [39]. Lignocellulose is comprised of three main components: cellulose, hemicellulose, and lignin. Cellulose, and to a lesser extent hemicellulose, are easily depolymerized through enzymatic degradation into their component sugar molecules [40]. Lignin, on the other hand, is a complex aromatic compound that is resistant to degradation by microorganisms [41]. Due to its recalcitrance, the lignin fraction is often discarded or burned to produce process heat or on-site electricity generation [42]. However, techno-economic analyses have identified the co-utilization of lignin as an important factor for the production of cost effective biofuels [43]. To aid in lignin degradation, lignin is often thermochemically depolymerized, resulting in toxic xenobiotics such as vanillic acid, catechol, carboxylic acids, and phenolic compounds that inhibit bioconversion [40, 41]. Research is currently underway to engineer a microbial strain, either through adaptive evolution or rational design, that is tolerant of and can degrade these xenobiotic compounds, thus increasing bioconversion productivity and yield [44].

My first objective in this thesis was to characterize the evolution of xenobiotic tolerance in a single bacterial strain after several generations of adaptive evolution to aromatic compounds. To evaluate this, we conducted adaptive evolution on *Rhodococcus opacus* PD630 (hereafter *R. opacus*) over several generations using various lignin-derived model compounds. *R. opacus* is a soil bacteria isolated from contaminated soil at a gasworks plant that has a naturally high tolerance for aromatic compounds [45]. Through adaptive evolution we increased the xenobiotic tolerance of *R. opacus* up to 1900% on several phenolic compounds and their mixtures. Using comparative

genomics of these adapted strains compared to the wild type strain, we identified genetic mutations and changes in gene expression levels that contribute to xenobiotic tolerance at the strain level.

1.5 Antibiotic Resistance in Gut Microbial Communities

Not all xenobiotic tolerance is useful from an anthropogenic viewpoint. Microbial resistance to antibiotics is quickly becoming a major public health problem [46]. Even though antibiotic resistance pre-dates mass production of antibiotics, widespread use of antibiotics has selected for an increased abundance of antibiotic resistance genes (ARGs) in environmental and human microbiomes [47-50]. Antibiotic resistance in bacteria is particularly troubling due to the fact that adaptive evolution is working against us, i.e. extensive treatment with antibiotics results in a higher abundance of ARGs and increased tolerance [51]. Thus, the only way to combat antibiotic resistance, besides continually designing new families of antibiotics, is to prevent the spread of plasmid-borne ARGs and develop ARG inhibitors that can rescue antibiotic potency [52]. Identifying ARGs before they are acquired and widely disseminated by pathogens is crucial to this effort and provides much needed time to develop ARG inhibitors [53, 54].

The second objective of my thesis was to examine the gut microbial community in humans and great apes along a gradient of Westernization. We compared the gut microbiome and resistome of wild chimpanzees and gorillas living in Nouabalé-Ndoki National Park in the Republic of the Congo, humans living just outside the park, and captive chimpanzees and gorillas from the St. Louis Zoo (USA) and the Lincoln Park Zoo (Chicago, USA). With this cohort we can investigate the differences in taxonomy and the abundance and richness of ARGs that occur in the gut microbiome in cohorts with and without antibiotic exposure. Finally, by using functional metagenomics we are able to identify novel ARGs that are present in these samples for further characterization.

Chapter 2

Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*

Collaboration Statement

This work was performed in a collaboration between William R. Henson, Tayte P. Campbell, Drew M. DeLorenzo, Yu Gao, Bertram Berla, Soo Ji Kim, Marcus Foston, Tae Seok Moon, and Gautam Dantas. M.F., T.S.M., and G.D. conceived the proect. W.R.H, T.P.C, D.M.D, Y.G., B.B., and S.J.K performed experiments. W.R.H., T.P.C, D.M.D, T.S.M, and GD analyzed the results. W.R.H, T.P.C, D.M.D, M.F, T.S.M, and G.D wrote the manuscript.

2.1 Abstract

Lignin utilization has been identified as a key factor in biorefinery profitability. However, lignin depolymerization generates heterogeneous aromatic mixtures that inhibit microbial growth and the conversion of lignocellulose to biochemicals. *Rhodococcus opacus* is a promising aromatic-catabolizing, oleaginous bacterium, but mechanisms for its aromatic tolerance and utilization remain undercharacterized. To better understand these mechanisms, we adaptively evolved *R. opacus* for improved utilization of 32 combinations of diverse aromatic compounds.

Evolved *R. opacus* mutants showed up to 1900% growth improvement in the utilization of phenol, guaiacol, 4-hydroxybenzoate, vanillate, and benzoate compared to the wild-type strain. Whole genome sequencing revealed several redox-related genes with mutations shared across multiple adapted mutants. PVHG6, the mutant with the most improved growth on a mixture of multiple aromatic compounds, showed 56% lower superoxide dismutase activity than the wild-type strain, suggesting that redox reactions are important for aromatic tolerance and utilization. Comparative transcriptomics revealed by-product detoxification pathways and five aromatic funneling pathways that were upregulated in response to specific aromatic compounds. Gene knockout experiments confirmed the two degradation routes of the β-ketoadipate pathway for five aromatic compounds. These results provide an improved understanding of aromatic bioconversion and facilitate development of *R. opacus* as a biorefinery host.

2.2 Introduction

Lignocellulosic biomass is a potential source of renewable bio-based fuels and chemicals. A large fraction of lignocellulose (~ 70–85%) is cellulose and hemicellulose, which can be depolymerized into monomeric sugars, such as glucose and xylose [55, 56]. The remaining 15–30% of lignocellulose is lignin, a highly cross-linked, heterogeneous, and recalcitrant aromatic polymer [57]. Microbial conversion of lignocellulose holds promise as a new way to produce a variety of novel products and petrochemical replacements (e.g., drop-in fuels) [58-67]. Current biomass pretreatment approaches, which aim to increase sugar extractability from lignocellulose, also release aromatics from lignin that negatively affect microbial product titers, yields, and productivities [41, 68]. Current processes separate the carbohydrate fraction from the lignin fraction, and the separated lignin is either discarded or burned for process heat or on-site electricity generation [69]. However, removal of lignin-derived inhibitors from sugar streams is still

challenging and costly, and techno-economic analyses have identified the co-utilization of lignin as an important factor in biorefinery profitability [43, 70]. Therefore, efforts are underway to develop microbial strains that can tolerate and convert inhibitory, lignin-derived compounds into value-added products [71-80].

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a promising bacterial strain for producing valuable products from lignocellulose. *R. opacus* is a gram-positive soil bacterium that has been shown to have an inherently high tolerance to aromatic compounds [45, 81]. Additionally, it accumulates high levels (up to ~78% as cell dry weight) of the biofuel precursor triacylglycerol (TAG) and demonstrates a moderately high growth rate [45, 82]. *R. opacus* can natively consume, or has been engineered to consume, hexose and pentose sugars present in lignocellulosic feedstocks [82, 83]. *R. opacus* can also consume aromatic compounds found in depolymerized lignin, such as phenol, 4-hydroxybenzoate, and vanillate [84-86], making it uniquely qualified to utilize all three primary depolymerized components of lignocellulose.

The co-consumption of mixtures of lignin-derived aromatic monomers is an underexplored area of study that is important for microbial lignin conversion strategies. The inherent heterogeneity of lignin is a major barrier to lignin valorization, and economical strategies require rapid conversion of complex aromatic mixtures to products [87-90]. Microbial degradation of diverse aromatic compounds is typically accomplished by biological funneling, where compounds are converted into common metabolites (e.g., catechol and proto-catechuate), prior to their catabolism via a central aromatic degradation pathway, such as the β-ketoadipate pathway [87, 91]. However, interactions between similarly structured compounds and different aromatic degradation pathways are not well understood in bacteria. Additionally, cross tolerance between lignin-derived aromatics has not been studied in *R. opacus*, although it is thought that low

concentrations of multiple inhibitory compounds can negatively affect bacterial growth more than high concentrations of a single inhibitory compound on its own [92].

The underlying mechanisms for aromatic tolerance and utilization in *R. opacus*, and microbes in general, are not well understood due to the complexity of the phenotype, which limits the use of rational strain engineering. One methodology for improving aromatic tolerance is through adaptive evolution, where evolutionary selection pressure is applied to the organism through serial subcultures [93, 94]. *R. opacus* has been demonstrated to be amenable to adaptation for improved growth in the presence of different inhibitors from lignocellulosic hydrolysates (e.g., furans, organic acids, and phenolics) [92], but the mechanistic basis for these improvements are unknown. Our previous work showed that *R. opacus* could be adapted for improved growth and lipid production on the lignin model compound phenol as a sole carbon source [86], and we identified the upregulation of aromatic degradation pathways and transporters as potential mechanisms of improvement in the adapted strains. However, phenol is only one of many potential aromatic compounds from lignin [88-90], and additional research is required to understand *R. opacus'* ability to adapt to mixtures of lignin-derived model compounds.

In this work, *R. opacus* was adaptively evolved on 32 combinations of six lignin model compounds (without added sugars) to study the capacity to improve its aromatic tolerance and utilization (Figure 2-1). Comparative genomic analysis of 35 improved adapted strains identified a high percentage of mutations in oxidation and reduction (redox) related genes, several of which were shared across multiple strains from different adaptation conditions. We demonstrated that the activity of one of these targeted enzymes, superoxide dismutase, was 56% lower in PVHG6 (a mutant adapted for improved growth on a mixture of multiple compounds) than the wild-type (WT) strain. This suggests that reactive oxygen species may be important for aromatic tolerance

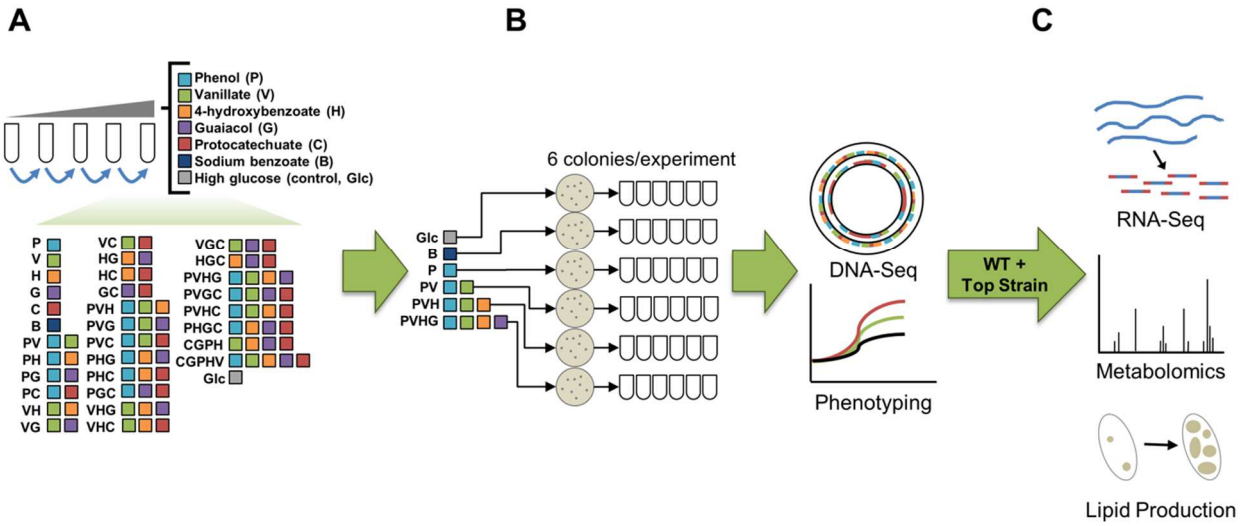


Figure 2-1: Experimental Approach. (A) *R. opacus* PD630 was serially passaged on lignin model compounds and high glucose (33 conditions). Six adapted cultures were selected for further characterization: four using phenol and up to three other lignin model compounds, one using sodium benzoate, and one using high glucose as carbon sources. (B) Six colonies from the selected adapted cultures were phenotyped in adaptation growth conditions to determine improvement compared to the WT strain. Strains with improved growth (OD_{600}) compared to the WT strain were characterized by whole genome sequencing and growth assays using mixtures of lignin model compounds. (C) The best-performing adapted strain (PVHG6) and the WT strain were further characterized using RNA-Seq and targeted metabolomic analyses to identify mechanisms for improved lignin model compound tolerance and utilization. Lipid titers were measured to characterize the adapted strain's capacity to convert lignin model compounds into lipids.

and utilization. To examine whether aromatics were consumed individually or concurrently, targeted metabolomics was performed, revealing that compounds were consumed in a distinct order and that the PVHG6 strain co-consumed lignin model compounds at faster rates than the WT strain. We performed RNA-Seq on PVHG6 and WT grown on individual lignin model compounds and their mixture (without added sugars) to elucidate genes and pathways related to aromatic catabolism. Transcriptomic analyses identified five distinct funneling pathways and two by-product detoxification pathways, each of which is activated by specific aromatic compounds, as well as a number of upregulated transporters. Additionally, by gene knockout experiments, we confirmed the degradation routes of five aromatic compounds and the importance of upregulated

transporters for growth on aromatic compounds. Lastly, to demonstrate that our improved aromatic tolerance and utilization phenotypes are compatible with production of valuable products from lignin, we measured lipid production and observed a 225% increase in lipid titer with the PVHG6 strain relative to the WT strain when grown on 2.5 g/L total aromatics as a sole carbon source. Together, these results advance our understanding of aromatic mixture tolerance and catabolism, which is a critical first step toward future fuel and chemical production from lignocellulose and establish *R. opacus* as a promising bioproduction strain.

2.3 Results

2.3.1 Adaptive evolution and characterization of adapted strains

To determine *R. opacus'* ability to improve its aromatic tolerance and consumption, we adaptively evolved it using increasingly higher concentrations of individual lignin model compounds and combinations of lignin model compounds (a total of 32 conditions without added sugars, Figure 2-1, Table 2-S1). For comparison to aromatic adaptation experiments, we also serially passaged *R. opacus* using high concentrations of glucose. For further analysis, we selected four adaptation experiments that used one to four different lignin model compounds that included phenol and other phenolic compounds as carbon sources, as well as two adaptation experiments that used sodium benzoate (a non-phenolic lignin model compound) and glucose as sole carbon sources. The adaptation experiments chosen for further characterization were:1) phenol (P); 2) phenol and vanillate (PV); 3) phenol, vanillate, and 4-hydroxybenzoate (PVH); 4) phenol, vanillate, 4-hydroxybenzoate, and guaiacol (PVHG); 5) sodium benzoate (B); and 6) high glucose. Six colonies from each of these six chosen adaptation experiments were isolated, and their growth was compared to that of the WT strain using their respective adaptation carbon source(s) (Figures

2-2, 2-S1). 35 of the 36 isolated strains had significantly higher cell densities (OD 600; up to1900% improvement) than the WT strain after 43 h of growth on the carbon sources used in their adaptation. Aromatic consumption pat-terns for adapted strains after 43 h of growth suggest that improved growth of these strains is due to both consumption of and tolerance to aromatic carbon sources (Figure 2-S2). These results demonstrate that *R. opacus* can be adaptively evolved to improve growth on different combinations of lignin model compounds.

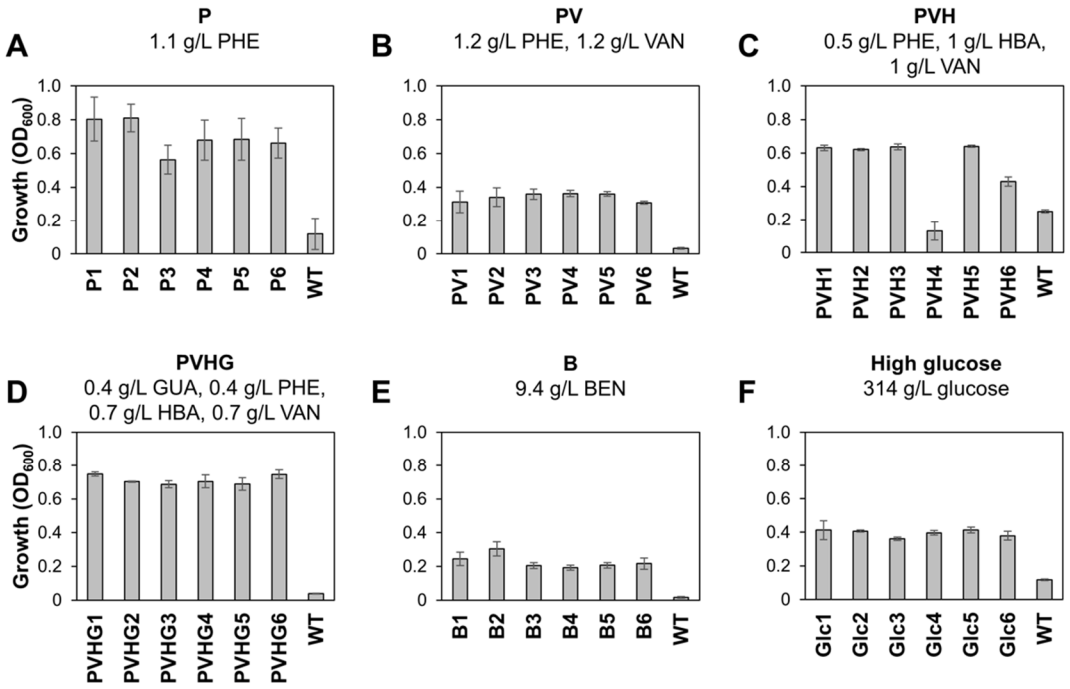


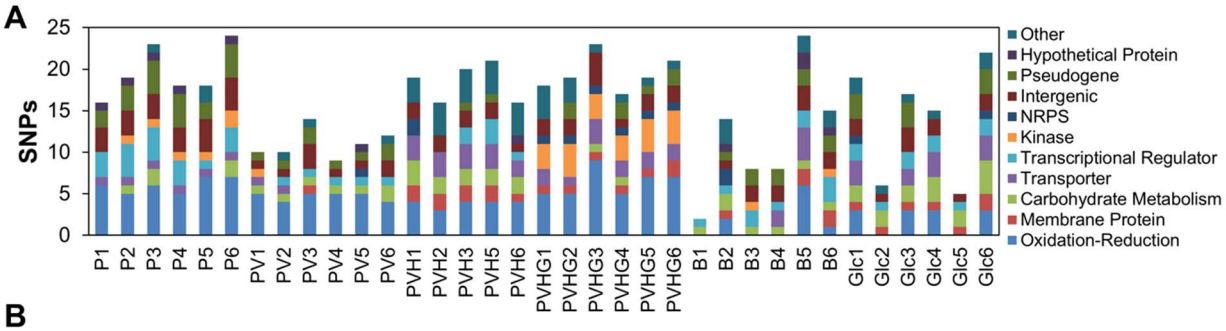
Figure 2-2: Growth of the WT strain and adapted strains using adaptation carbon sources. Bars represent OD_{600} values measured at 43 h of growth in minimal medium A supplemented with the listed carbon sources and 1 g/L ammonium sulfate as the nitrogen source. (A) The PHE adapted strains (P1-6) compared to the WT strain using 1.1 g/L PHE. (B) The PHE and VAN adapted strains (PV1-6) compared to the WT strain using 1.2 g/L PHE and 1.2 g/L VAN. (C) The PHE, VAN, and HBA adapted strains (PVH1-6) compared to the WT strain using 0.5 g/L PHE, 1 g/L HBA, and 1 g/L VAN. (D) The PHE, VAN, HBA, and GUA adapted strains (PVHG1-6) compared to the WT strain using 0.4 g/L GUA, 0.4 g/L PHE, 0.7 g/L HBA, and 0.7 g/L VAN. (E) The BEN adapted strains (B1-6) compared to the WT strain using 9.4 g/L BEN. (F) The high glucose adapted strains (Glc1-6) compared to the WT strain using 314 g/L glucose. Bars represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

2.3.2 Comparative genomics of adapted strains

To identify possible causative mutations for adaption, we examined the 35 isolated strains with improved growth (Figure 2-2) by whole genome sequencing (> 38× coverage per strain). Across all 35 strains, genomic analysis revealed a total of 255 distinct SNPs and 6 insertion/deletions (INDELs) distributed over the chromosome and 6 of the 9 endogenous plasmids [95] (Figure 2-S3, Tables 2-S2–2-S4). Based on sequencing depth, we observed major deletions in Plasmid 1 for 11 strains from the PVHG, B, and high glucose adaptation experiments, and we observed complete loss of Plasmid 2 in 25 strains from the P, PV, PVH, PVHG, and B adaptation experiments (Figure 2-S4). Plasmids 1 and 2 are large plasmids with sizes of 172,218 bp and 97,588 bp, respectively. Growth improvements could result from either deletion of burdensome genes, reduced replication costs for maintaining large plasmids, or a combination of both [96, 97].

To identify potential common patterns of genetic changes between strains and adaptation conditions, we classified genes affected by mutations using Gene Ontology (GO) terms [98] and then further grouped GO terms into broader functional categories (Figure 2-3A). Interestingly, many genes involved in redox reactions were affected by SNPs after adaptation in phenolic compound mixtures (Figure 2-3A). On average, 36% of the non-synonymous SNPs in strains adapted using phenolic compounds, compared to only 16% in strains adapted using non-phenolic carbon sources (e.g., BEN and glucose), were found in genes related to redox reactions. These results suggest that the selection pressure applied by adaptive evolution targeted redox reactions to improve tolerance and utilization of phenolic compounds.

We identified seven genes that were affected by non-synonymous SNPs across multiple aromatic adaptation experiments (Figure 2-3B, Table 2-S5). Surprisingly, we did not observe any shared genes whose annotation is related to aromatic degradation. Of the seven genes, three genes



B

Strains	Gene ID	Annotation	Location	Base Mean Expression
All P, All PV, PVH6, All PVHG	RS01890	superoxide dismutase	Chromosome	32,010
All B, All PVH	RS08140	phosphoenolpyruvate carboxykinase	Chromosome	30,198
P2, P3, P6, All PV	RS10095	class II fructose-bisphosphate aldolase	Chromosome	5,291
All P, All PV, All PVH, All PVHG	RS14825	cytochrome ubiquinol oxidase subunit I	Chromosome	23,629
PV3, All PVHG	RS32480	phosphotransferase	Chromosome	285
P5, PV3, PVH3, PVH5, B6	RS38570	FAD-binding oxidoreductase	Plasmid 1	60
P2, P3, P4, P5, P6, All PVHG	RS39015	serine/threonine-protein kinase PknK	Plasmid 3	354

Figure 2-3: Comparative genomics of lignin model compound adapted strains. (A) Categorization of genes with non-synonymous SNPs or intergenic SNPs that could affect gene expression using Gene Ontology (GO) terms organized by strain. Strain names indicate the adaptation condition from which each strain originated: P = phenol; PV = phenol and vanillate; PVH = phenol, vanillate, and 4-hydroxybenzoate; PVHG = phenol, vanillate, 4-hydroxybenzoate, and guaiacol; B = sodium benzoate; and Glc = glucose. (B) Genes with non-synonymous SNPs or intergenic SNPs that could affect gene expression and that were found across multiple aromatic adaptation experiments. Putative annotations are from NCBI reference sequence NZ_CP003949.1. "All" indicates that all sequenced strains isolated from the indicated adaptation condition contain either non-synonymous SNPs or intergenic SNPs that could affect gene expression. Base Mean Expression is the average of the normalized counts from WT and PVHG6 transcriptomic data from all growth conditions.

are annotated to be involved in redox reactions: RS14825 (cytochrome ubiquinol oxidase subunit I), RS01890 (manganese-dependent superoxide dismutase [SOD]), and RS38570 (a FAD-binding oxidoreductase). Non-synonymous or intergenic SNPs affecting cytochrome ubiquinol oxidase and superoxide dismutase were found in at least one strain from all four adaptation experiments containing phenolic compounds, but not in any strains from the B and high glucose adaptation experiments. The previous adaptation experiment in *R. opacus* using phenol [86] also led to a mutation in cytochrome ubiquinol oxidase. Furthermore, two genes involved in gluconeogenesis

had non-synonymous SNPs in multiple adaptation lineages: RS08140 (phosphoenolpyruvate [PEP] carboxykinase) and RS10095 (class II fructose-bisphosphate aldolase). Additionally, two genes related to phosphorylation had non-synonymous SNPs: RS32480 (phosphotransferase) and RS39015 (serine/threonine protein kinase).

To investigate the consequence of non-synonymous SNPs in the cytochrome ubiquinol oxidase and SOD genes further, we modeled the mutational sensitivity (i.e., the probability of the mutation having a phenotypic effect) using Phyre2 [99] and SuSPect [100]. The modified residues in the adapted strains for both genes had mutational sensitivities ranging from medium to very high for the modified amino acids, suggesting that these SNPs could have a phenotypic effect (either positive or negative; Tables 2-S6, 2-S7) [100]. To verify the modeling results for superoxide dismutase, we measured the SOD activity using cell lysates from the WT strain and one of our multi-compound adapted strains with a non-synonymous SNP in SOD (PVHG6). The adapted PVHG6 strain showed a 56% lower SOD activity compared to that of the WT strain (Figure 2-S5). This lower activity suggests that the adapted strain is likely to reduce fewer superoxide radicals than the WT strain, which could contribute to its improved growth in the mixture of lignin model compounds through their facilitated oxidation [101, 102].

2.3.3 Characterization of adapted strains on lignin model compound mixtures

Efficient microbial lignin valorization requires the utilization of complex aromatic compound mixtures. Thus, we sought to determine the performance of our aromatic adapted strains in mixtures of lignin model compounds compared to the WT strain. We hypothesized that the adapted strains' improved tolerance phenotype towards a subset of five chosen lignin model compounds (one to four compounds) could be extended to a mixture of all five compounds. To test this hypothesis, we grew one strain from each of the P, PV, PVH, PVHG, and B adaptation

experiments on a mixture of five aromatics (PHE, GUA, HBA, VAN, and BEN). At 2.5 g/L total aromatics (0.5 g/L of each lignin model com-pound), all five adapted strains had significantly higher growth rates than the WT strain ($P < 0.009$ for all adapted strains; Figure 2-4A, Table 2-S8). P1, PV1, and PVHG6 strains reached a higher OD_{600} than the WT strain after 100 h of growth ($P < 0.004$ for all three strains), while PVH5 and B2 did not reach significantly higher OD_{600} values (Table 2-S8). PVHG6 reached the highest OD_{600} among all strains after 70 h of growth, which was 96% higher than the WT strain ($P = 3.8 \times 10{-}6$). At 3.0 g/L total aromatics (0.6 g/L of each aromatic compound), all strains had significantly higher growth rates than the WT strain ($P < 0.03$ for all adapted strains; Figure 2-S6, Table 2-S9). P1, PV1, and PVHG6 also reached 75–85% higher OD_{600} values than the WT strain after 100 h of growth ($P < 0.03$ for all three strains), while PVH5 and B2 did not reach significantly higher OD_{600} values (Figure 2-S6, Table 2-S9). These results demonstrate that some adapted strains (P1, PV1, and PVHG6) have improved growth on a mixture of aromatics, some of which were not used during their adaptation. For the three best-performing strains (P1, PV1, and PVHG6), the only exclusively shared gene which was affected by a SNP was SOD (Figure 2-3B), which further supports that SOD may be related to improved growth on lignin model compounds. Because PVHG6 showed the best growth performance at both 2.5 and 3.0 g/L total aromatics, it was chosen for further characterization.

2.3.4 Aromatic consumption in a lignin model compound mixture by PVHG6

Next, we sought to determine if all lignin model compounds in the five-compound mixture were used for growth and if PVHG6 consumed aromatics in a similar manner to that of the WT strain. We used a targeted metabolomic approach to measure the concentration of each individual aromatic compound in the mixture (0.5 g/L of PHE, GUA, HBA, VAN, and BEN; 2.5 g/L total aromatics; Figure 2-4B). We observed that both WT and PVHG6 strains consumed BEN and HBA

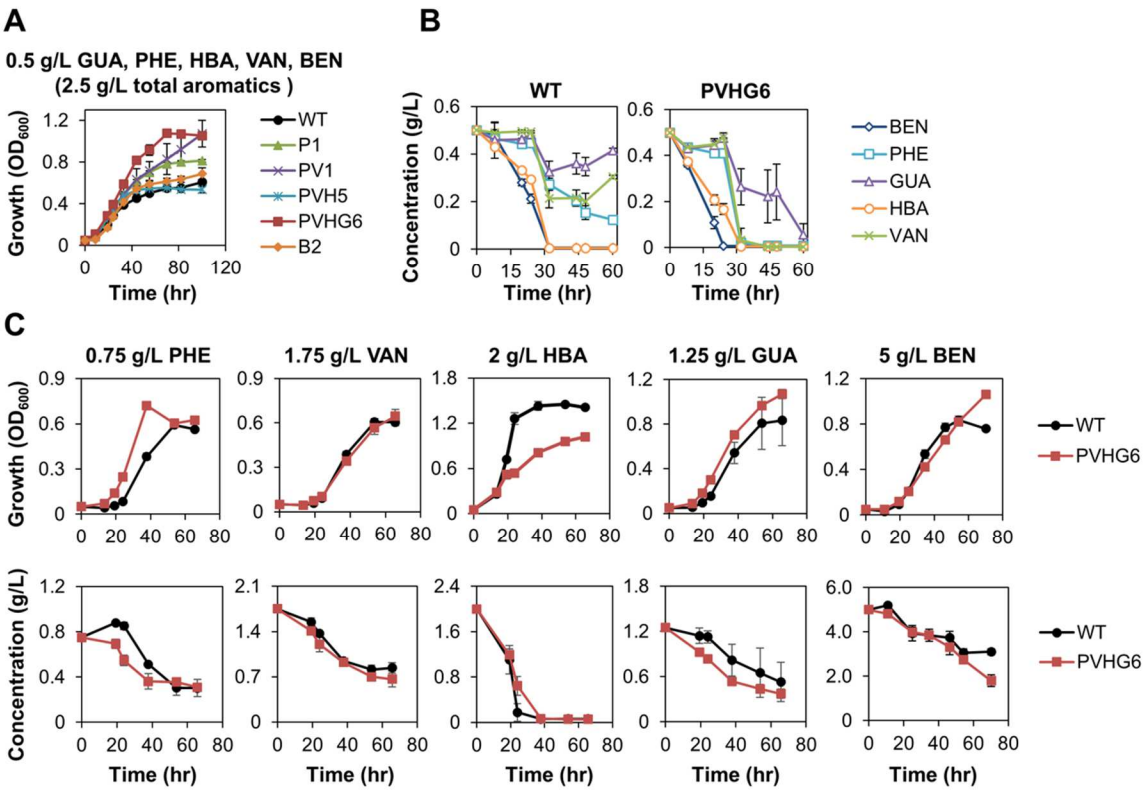


Figure 2-4: Growth and consumption of lignin model compounds by WT and adapted strains. Cells were grown in minimal medium A supplemented with the listed carbon sources and 1 g/L ammonium sulfate as the nitrogen source. (A) Growth of one strain from each aromatic adaptation experiment and the WT strain in a mixture of lignin model compounds (0.5 g/L each of GUA, PHE, HBA, VAN, and BEN; 2.5 g/L total aromatics). (B) Aromatic consumption profiles of WT and PVHG6 strains in the same growth condition as A. Aromatic concentrations were measured by derivatizing culture supernatants using methyl chloroformate followed by GC-MS-FID (see Section 2.5 for details). (C) Growth and consumption of individual aromatic compounds as a sole carbon source by WT and PVHG6 strains (aromatic concentrations listed above plots). Aromatic concentrations were estimated using UV absorbance (see Section 2.5 for details). Note that the y-axis scales for growth and concentration plots vary between carbon sources. For all plots, points represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

before PHE, GUA, and VAN. Both strains fully consumed BEN and HBA within 24–32 h, while consumption of PHE, GUA, and VAN began after 24 h for both strains. Within 60 h, all five compounds were almost fully consumed by the PVHG6 strain, but substantial amounts of PHE (25%), GUA (83%), and VAN (61%) remained in the WT cultures at 60 h. Overall, PVHG6 had

a 27% higher total aromatic consumption rate than the WT strain (52 mg/L/h for PVHG6 vs. 41 mg/L/h for the WT strain from 0 to 44 h, P = 0.0024). Additionally, it was found that PVHG6 also consumed a mixture containing 3 g/L glucose and all five aromatic compounds (0.6 g/L of PHE, GUA, HBA, VAN, and BEN; 3 g/L total aromatics) as carbon sources, and it completely utilized all aromatics after 60 h of growth (Figure 2-S7).

2.3.5 Consumption of individual aromatic compounds by PVHG6

To determine whether the improved growth phenotype of PVHG6 on a mixture of lignin model compounds extended to individual compounds, we measured growth and aromatic concentrations using six different concentrations of PHE, VAN, HBA, GUA, and BEN as a sole carbon source (Figures 2-4C, 2-S8–2-S13). Aromatic concentrations were measured using UV absorbance. From these experiments, we calculated the half maximal inhibitory concentration (IC_{50}) using OD_{600} values at 54 h. We only observed a significant increase in the IC_{50} value of PVHG6 compared to that of the WT strain when they were grown in PHE (17%, P = 0.003) and GUA (16%, P = 0.008) (Table 2-S10). The IC_{50} values of PVHG6 for HBA, VAN, and BEN were unchanged or slightly decreased (Table 2-S10). When 1 g/L glucose was used as a sole carbon source, no difference in growth was observed between the WT and PVHG6 strains (Figure 2-S14), implying that glucose metabolism was not affected by mutations during the long adaptation process in the absence of glucose.

2.3.6 Transcriptomic analysis of WT and PVHG6 using lignin model compounds

We had previously observed that changes in gene and pathway expression better explained improved phenol utilization phenotypes in evolved *R. opacus* strains, compared to the effect of genetic changes like SNPs and INDELs in coding regions [86]. Accordingly, we sought to

determine the potential role of transcriptional differences in the improved tolerance and utilization of a larger panel of aromatics by the evolved PVHG6 strain. We performed whole cell gene expression profiling via RNA-Seq on the WT and PVHG6 strains grown in a mixture of all five lignin model compounds (0.5 g/L of each lignin model compound, 2.5 g/L total aromatics), or in 1 g/L glucose, to identify differentially regulated genes between strains and carbon sources. We also performed RNA-Seq using 0.5 g/L of each of the five lignin model compounds individually (PHE, GUA, HBA, VAN, or BEN) to determine compound-specific degradation pathways for each compound. For the 1 g/L glucose (Glc) and the mixture of five lignin model compounds (hereafter, "mixture"), we analyzed the transcriptome at both the early and mid-exponential phases (20 h and 32 h for the mixture; 10 h and 13 h for Glc) since we had observed differential consumption rates of lignin model compounds in our targeted metabolomic analysis (Figure 2-4B).

To investigate differences in expression profiles between WT and PVHG6, we first analyzed all transcriptomic data for both strains using principal component analysis (Figure 2-S15). PVHG6 and WT were clearly separated along the first principal component with values higher than 60 and lower than −50, respectively, while the second principal component generally separated transcriptional profiles by growth conditions. Permutational ANOVA (PERMANOVA) indicated that transcriptomic profiles between strains and between growth conditions were significantly different ($P = 0.001$, both cases). Based on our genome sequencing data, which was corroborated by our transcriptomic data (Figure 2-S4), many adapted strains, including PVHG6, underwent plasmid deletions and plasmid loss. We repeated the principal component analysis with Plasmid 1 and 2 data removed, and while some growth conditions led to closer visual clustering

of WT and PVHG6, the strains were still significantly different based on PERMANOVA (P = 0.002; Figure 2-S16).

We then compared the top 30 differentially upregulated genes between both strains grown in the mixture at both time points (Figure 2-S17).The most upregulated genes in PVHG6 relative to the WT strain in the mixture at both time points were RS27075 (a hypothetical protein, 41-fold upregulated) and RS13410 (a type VII secretion-associated protein, 40-fold upregulated), respectively. Five of the top 30 upregulated genes are annotated as ABC transporters or ABC transporter associated proteins, suggesting that these ABC transporters might control uptake or export rates of aromatic compounds. Two genes annotated as acetyl-CoA acetyltransferases (RS34095 and RS34105) were also in the top 30 list. Both proteins share high positive amino acid sequence similarity to 3-oxoadipyl-CoA thiolases in *R. opacus* and *R. jostii* RHA1 (52–76% positive identity, Table 2-S11). The final step in the β-ketoadipate pathway is catalyzed by 3-oxoadipyl-CoA thiolase, converting 3-oxo-adipyl-CoA into succinyl-CoA and acetyl-CoA. RS34095 and RS34105 may encode this enzyme, acting as additional gene copies and increasing the metabolic flux through the β-ketoadipate pathway in PVHG6 when the aromatic mixture is used.

To identify putative funneling pathway enzymes for each lignin model compound, we searched for the most strongly upregulated enzymes when both strains were grown in the individual aromatic compounds and the mixture, compared to the glucose condition (Figure 2-5A). These selected proteins were then compared to homologous enzymes in a closely-related strain, *R. jostii* RHA1, which has a characterized proteome [103] (Table 2-S12). Using this process, we identified funneling pathway clusters for each lignin model compound: two two-component phenol hydroxylases (RS31555–RS31560 and RS30765–RS30770), a putative vanillate

demethylase (RS02665–RS02675), a putative 4-hydroxybenzoate monooxygenase (RS31675), two putative cytochrome P450 clusters upregulated in response to GUA (RS30780–RS30785 [GUA degradation cluster #1] and RS21470–RS21485 [GUA degradation cluster #2]), and a putative benzoate degradation cluster (RS30790–RS30805). Many of these funneling pathway enzymes have high sequence similarity to biochemically characterized pathway enzymes in other *Rhodococcus* strains, including *R. opacus* 1CP (Table 2-S12) [104, 105]. Each of these genes or gene clusters converts lignin model compounds to either catechol (CAT) or protocatechuate (PCA) for subsequent degradation by the β-ketoadipate pathway (Figure 2-6B). All of these genes were upregulated in both the individual aromatic condition and the mixture compared to glucose, except GUA degradation cluster #2 which was upregulated in GUA but not upregulated in the mixture compared to glucose (Figure 2-5A). This differential expression suggests transcriptional regulation (i.e., repression) of this funneling pathway in response to the presence of other lignin model compounds. When the expression of the identified funneling pathways in PVHG6 was compared to the WT strain in the mixture, most genes at both time points were not differentially expressed with the exception of the benzoate degradation pathway at the second time point, which was 29–95-fold downregulated (adjusted p-values $< 7 \times 10{-}29$). Our metabolomic data (Figure 2-4B) showed that PVHG6 completely consumed BEN after 24 h (compared to 32 h for WT), which suggests that the BEN funneling pathway genes were downregulated in PVHG6 in response to complete consumption of BEN. For other funneling pathways, no genes were more than 2.6-fold upregulated or downregulated in PVHG6 compared to the WT strain at either time point in the mixture.

The funneling pathways convert aromatic compounds into two intermediates, CAT or PCA, for degradation by the β-ketoadipate pathway (Figure 2-6). Previous analyses of

Rhodococcus strains suggest that VAN and HBA are converted to PCA, while GUA, BEN, and PHE are converted to CAT [106]. For both WT and PVHG6, we observed upregulation of β-ketoadipate gene cluster #1 (RS25340–RS25375) compared to the glucose condition for all aromatic carbon sources (Figure 2-6A). This gene cluster includes genes required to convert both CAT and PCA into central tricarboxylic acid (TCA) cycle metabolites. In both strains, β-ketoadipate gene cluster #2 (RS31565–RS31575) was upregulated in PHE and the mixture (up to 5000-fold), but it was upregulated only up to 3-fold in other lignin model compounds compared to the glucose condition. This observed upregulation could be related to the upstream phenol hydroxylase (RS31555–RS31560 in the phenol funneling pathway; Figure 2-5A) that is highly upregulated in PHE compared to the glucose condition. We also observed upregulation of β-ketoadipate gene cluster #3 (RS30720–RS30745) during growth on all aromatic carbon sources. In both strains, genes in this cluster, which is typically associated with degradation of CAT intermediates, were highly upregulated (142–779-fold) by CAT branch compounds (GUA, BEN, and PHE) relative to the glucose condition, but they were only moderately upregulated (9–93-fold) by PCA branch compounds (VAN and HBA) relative to the glucose condition. Interestingly, such a branch-dependence was also observed in regulation of some funneling enzymes (Figure 2-5A). For example, a phenol hydroxylase (RS30765–RS30770) was also highly upregulated in GUA (121–477-fold), but minimally upregulated by PCA precursors VAN and HBA (up to 5.9-fold), compared to the glucose condition in both strains. The putative vanillate demethylase (RS02665–RS02675) was also upregulated in HBA (114–295-fold), but minimally upregulated by CAT precursors GUA and BEN (1.4–2.2-fold) in both WT and PVHG6, compared to the glucose condition. When compared to the WT strain in the mixture, PVHG6 did not show substantial upregulation of the β-ketoadipate clusters (at most 2.5-fold). This result differs from that of

previously characterized *R. opacus* mutants that were adapted using a single compound (PHE) and that showed large differential upregulation of a number of phenol degradation genes in PHE compared to the WT strain [86].

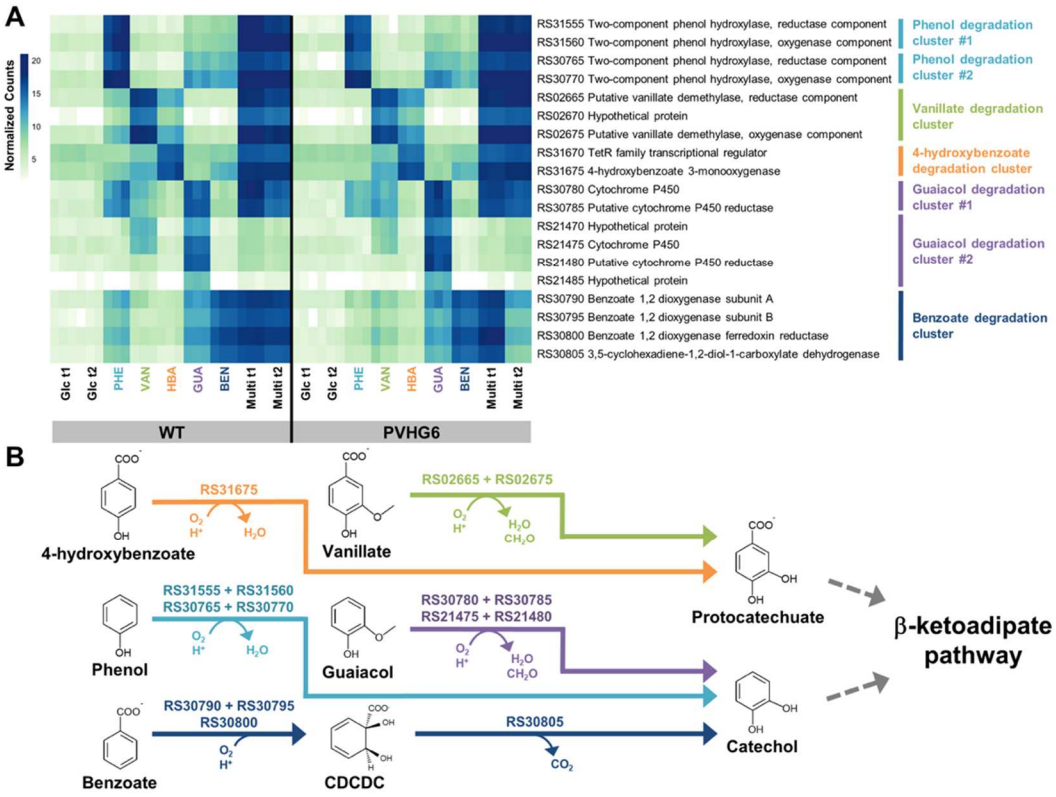


Figure 2-5: Activation of lignin model compound funneling pathways in *R. opacus*. (A) Heat map of normalized expression counts for upregulated aromatic degradation genes that convert (i.e. "funnel") lignin model compounds to the β-ketoadipate pathway during growth on individual lignin model compounds, a mixture of lignin model compounds, and glucose. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 g/L 4-hydroxybenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 h for Glc; 20 h for Multi), while t_2 represents cells harvested at mid-exponential phase (13 h for Glc; 32 for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 h for WT PHE, 21 h for PVHG6 PHE, 19 h for WT and PVHG6 GUA, 11 h for WT and PVHG6 HBA, 12 h for WT and PVHG6 BEN, 24 h for WT and PVHG6 VAN). Cells were grown in minimal medium A. Each square on the heat map represents a biological replicate. (B) Pathway map showing genes involved in lignin model compound funneling pathways in *R. opacus*. Gene codes are from the NCBI reference sequence NZ_CP003949.1. CDCDC = cis-1,2-dihydroxy-cyclohexa-3,5-diene-1-carboxylate.

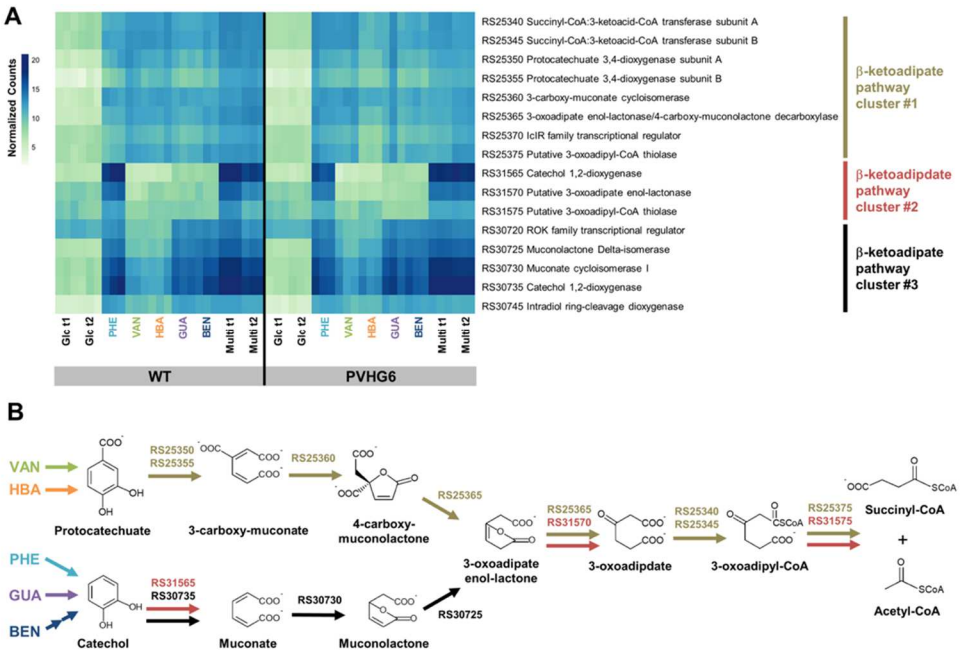


Figure 2-6: Activation of β-ketoadipate pathway gene clusters by lignin model compounds in *R. opacus*. (A) Heat map of normalized gene expression counts for β-ketoadipate pathway gene clusters during growth on individual lignin model compounds, a mixture of lignin model compounds, and glucose. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 0.5 g/L 4-hydroxybenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 h for Glc; 20 h for Multi), while t_2 represents cells harvested at mid-exponential phase (13 h for Glc; 32 h for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 h for WT PHE, 21 h for PVHG6 PHE 19 h for WT and PVHG6 GUA, 11 h for WT and PVHG6 HBA, 12 h for WT and PVHG6 BEN, 24 h for WT and PVHG6 VAN). Cells were grown in minimal medium A. Each heat map square represents a biological replicate. (B) Pathway map showing genes involved in β-ketoadipate pathway gene clusters in *R. opacus*. Gene codes are from the NCBI reference sequence NZ_CP003949.1. (C) Difference between initial OD_{600} and final OD_{600} after 24 h of growth for the RS30730 knockout (ΔCAT) strain and the WT strain using different carbon sources. (D) Difference between initial OD_{600} and final OD_{600} after 24 h of growth for the RS25360 knockout (ΔPCA) strain and the WT strain using different carbon sources. For C and D, the WT, ΔCAT, and ΔPCA strains were initially cultured in 2 mL of minimal medium B containing 1 g/L ammonium sulfate, 4 g/L glucose, and either no aromatic compound, 0.2 g/L PHE, 0.25 g/L GUA, 0.5 g/L HBA, 0.25 g/L VAN, or 0.5 g/L BEN. Cells were then centrifuged, washed with minimal medium B containing no carbon source, and re-suspended to an initial OD_{600} of 0.2 in 10 mL of minimal medium B containing 1 g/L ammonium sulfate and either 1 g/L glucose, 0.4 g/L PHE, 0.5 g/L GUA, 1.0 g/L HBA, 0.5 g/L VAN, or 1 g/L BEN as sold carbon sources. Asterisk (*) indicates that the difference between initial and final OD_{600} values is less than one standard deviation. Bars represent the average of three biological replicates and error bars represent one standard deviation.

To confirm whether the five tested aromatic compounds (PHE, GUA, HBA, VAN, and BEN) were catabolized via the CAT or PCA branch of the β-ketoadipate pathway, critical genes from both branches were knocked out, and growth assays were performed. β-ketoadipate gene cluster #3, which contains the CAT degradation pathway, was knocked out by a disruption of RS30730 (cis,cis-muconate cycloisomerase; ΔCAT), while β-ketoadipate gene cluster #1, which contains the PCA degradation pathway, was knocked out by a disruption of RS25360 (3-carboxy-cis,cis-muconate cycloisomerase; ΔPCA). The knockout of the PCA degradation pathway was complicated by the fact that other central β-ketoadipate pathway genes required for CAT degradation are located further downstream in β-ketoadipate gene cluster #1 (Figure 2-6). To ensure transcription of those downstream genes, a constitutive promoter was integrated into the genome. The ΔCAT strain was unable to grow using PHE, GUA, and BEN as sole carbon sources, while the ΔPCA strain was unable to grow using HBA and VAN as sole carbon sources (Figures 2-6C,D). Together, these results confirm that PHE, GUA, and BEN are funneled through the CAT branch, and HBA and VAN are funneled through the PCA branch of the β-ketoadipate pathway in *R. opacus*.

Transporter expression control may be related to aromatic tolerance mechanisms. The putative shikimate transporter (RS31355), which was significantly upregulated in other phenol-adapted *R. opacus* strains [86], was also upregulated in PHE (281–299-fold) and the mixture (87–295-fold) for both WT and PVHG6 relative to the glucose condition at both time points. An MFS transporter (RS33590) was moderately upregulated in response to PCA precursors VAN (137–193-fold) and HBA (10–15-fold), and the mixture (45–115-fold), but was minimally upregulated by CAT precursors PHE, GUA, and BEN (1.0–3.7-fold) in WT and PVHG6 at both time points. Another MFS transporter (RS30810) located adjacent to the BEN degradation cluster was

upregulated in BEN (1330–1720-fold) and the mixture (958–2620-fold) for both WT and PVHG6 at the first timepoint. At the second time point, the MFS transporter was down-regulated in PVHG6 compared to the WT strain in a similar manner to the BEN degradation cluster (Figure 2-5A). The MFS transporter RS30810 was also moderately upregulated by CAT precursor GUA (57–62-fold), but only minimally upregulated by PCA precursors VAN (1.6–7.5-fold) and HBA (1.1–8.7-fold) in both strains compared to the glucose condition. The branch-dependent upregulation of MFS transporters RS33590 and RS30810 is similar to what was observed for the β-ketoadipate pathway (Figure 2-6). Overall, we observed multiple transporters whose expression was responsive to lignin model compounds.

To explore the role of transporters in aromatic tolerance and utilization, we compared the growth of three putative transporter knockout mutants (RS31355, RS33590, and RS30810) to that of the WT strain using different carbon sources (Figure 2-S18). The growth of these knockout mutants and the WT strain were compared using 0.75 g/L PHE, 1.25 g/L GUA, 1.75 g/L VAN, 2 g/L HBA, 5 g/L BEN, and the combined mixture (2.5 g/L total aromatics) as sole carbon sources. The putative shikimate transporter (RS31355) knockout mutant had lower cell densities (OD_{600}) after 48 h than the WT strain using PHE and the mixture as sole carbon sources, which matches previous reports of RS31355 having a role in PHE transport [86]. The putative vanillate transporter (RS33590) knockout mutant had lower OD_{600} values than the WT strain using VAN, PHE, and the mixture as sole carbon sources. The putative benzoate transporter (RS30810) knockout mutant had lower OD_{600} values than the WT strain using BEN, VAN, GUA, PHE, and the mixture as sole carbon sources (Figure 2-S18). None of the transporter knockout mutants exhibited growth impairments on HBA (Figure 2-S18). The impaired growth of the putative aromatic transporter knockout mutants on some compounds suggests transporter specificity and that the influx or efflux

of aromatic compounds plays an important role in aromatic tolerance and utilization (Figure 2-S18).

The consumption of aromatics can lead to the formation of by-products. We observed upregulation of gene clusters involved in one carbon compound (C1) metabolism during growth on aromatics (Figure 2-S19). Catabolism of VAN and GUA requires a demethylation step that releases formaldehyde during conversion to PCA and CAT, respectively [107-109] (Figures 2-5, 2-6). In VAN and GUA growth conditions, we observed the upregulation of a mycothiol-dependent pathway that converts formaldehyde to CO_2, and this pathway is found in *Rhodococcus* strains and other actinomycetes [110, 111] (Figure 2-S19). This pathway (RS31920–RS31930) was upregulated in VAN and GUA (21–105-fold), but expression varied in the mixture from no significant upregulation to 26-fold upregulation for both strains compared to the glucose control. Another pathway for formaldehyde degradation is the ribulose mono-phosphate (RuMP) pathway, which converts formaldehyde and ribulose-5-phosphate to fructose-6-phosphate for consumption via the pentose phosphate pathway [112]. The RuMP pathway (RS21455–RS21460) was upregulated in VAN and GUA (450–2600-fold) compared to the glucose condition in both the WT strain and PVHG6. In the mixture, the RuMP pathway was upregulated in the WT strain (23–135-fold) and PVHG6 (3–172-fold) relative to the glucose condition. Genes in the RuMP pathway are also located near the pentose phosphate pathway operon, which was upregulated in a similar pattern to the RuMP pathway (Figure 2-S19). Together, these results suggest that *R. opacus* has multiple formaldehyde degradation pathways which are activated only when they are necessary (e.g., when VAN or GUA is catabolized, generating the toxic by-product).

To better understand the effect of mutations on the improved growth phenotype of PVHG6, we compared the expression of genes with mutations found in PVHG6 as well as genes with

mutations found in multiple adaptation conditions (Figures 2-S20, 2-S21). Of the genes with mutations in PVHG6, SOD (RS01890), non-ribosomal peptide synthetase (RS13070), and cytochrome ubiquinol oxidase (RS14825) were highly expressed (> 500 normalized counts) in both the WT and PVHG6 strains across all tested growth conditions, including the high glucose condition (Figure 2-S20). For genes with mutations found across multiple adaptation conditions (Figure 2-3B, Table 2-S5), SOD (RS01890), cytochrome ubiquinol oxidase (RS14825), phosphoenolpyruvate carboxykinase (RS08140), and fructose bisphosphate aldolase (RS10095) were highly expressed (> 500 normalized counts, Figure 2-S21). These results suggest that mutations could play a role in phenotypic differences between WT and PVHG6 strains, possibly by changing specific enzyme activities (e.g., SOD, Figure 2-S5), instead of changing expression levels of highly expressed genes (Figures 2-S20, 2-S21).

2.3.7 Lipid production using lignin model compounds

To determine the effect of the adaptation on the ability of *R. opacus* to convert lignin model compounds into a bioproduct (e.g., lipids), we compared total lipid production between WT and PVHG6 grown in 2.5 g/L of lignin model compounds (0.5 g/L each of PHE, GUA, HBA, VAN, and BEN) in minimal medium A in a low nitrogen condition (Figure 2-7). PVHG6 reached a 50% higher cell density (OD_{600}) than the WT strain after 72 h, and it accumulated 44% of its cell dry weight as lipids, compared to only 26% by WT (Figure 2-7). After 72 h, the adapted strain reached a lipid titer of 0.13 g/L using lignin model compounds as a sole carbon source, which is a 225% increase compared to the WT strain (Figure 2-7B). These results suggest that adaptive evolution can lead to significant improvements in lipid production by increasing lignin model compound tolerance and consumption.

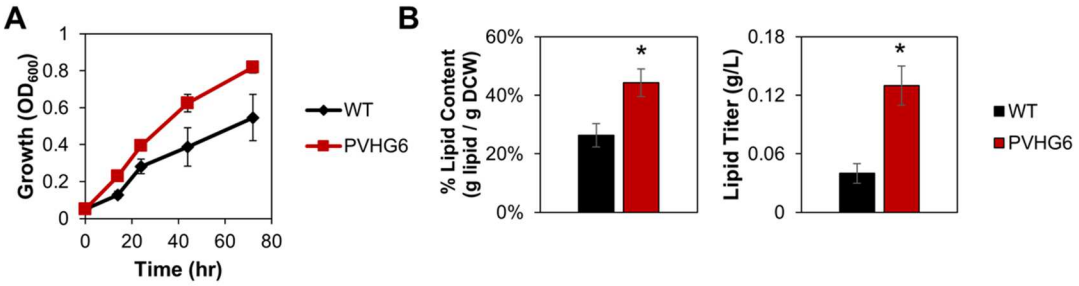


Figure 2-7: Conversion of lignin model compounds to lipids. (A) Growth of WT and PVHG6 strains in low nitrogen conditions using minimal medium A with 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics) as carbon sources and 0.125 g/L ammonium sulfate as the nitrogen source. (B) Lipid content and lipid titer of WT and PVHG6 strains after 72 h of growth in the same growth conditions as A. For all plots, points and bars represent the average of three biological replicates and error bars represent one standard deviation. Asterisk (*) indicates that the difference between WT and PVHG6 values is statistically significant ($P < 0.05$, one mean, two-tailed Student's *t*-test).

2.4 Discussion

Microbial tolerance and consumption of aromatic mixtures is a complex, understudied phenotype that involves numerous interactions between many genes and pathways. An improved understanding of this complex phenotype and its underlying mechanisms is critical for efficient lignin valorization. In this work, we investigated lignin model compound tolerance and utilization by combining adaptive evolution and multi-omic approaches. We demonstrated that *R. opacus* can be adaptively evolved to improve consumption of aromatic mixtures, and we identified shared mutations in adapted strains from multiple adaptation experiments that could affect aromatic tolerance and utilization. We observed that the adapted strain PVHG6 had higher aromatic consumption rates than the WT strain in a mixture of lignin model compounds, and we identified five upregulated funneling pathways and two by-product detoxification pathways that were responsive to specific lignin model compounds. Furthermore, our results demonstrate that enzymatic activities which affect levels of intracellular molecular oxygen and superoxide radicals

likely play an important role in aromatic tolerance and utilization. Overall, this work highlights the value of combining adaptive evolution with a multi-omic approach to provide insights into microbial aromatic tolerance and utilization, to create improved strains for aromatic mixture consumption, and to identify novel targets for engineering *R. opacus* strains toward lignin valorization.

The ability of *R. opacus* to successfully adapt to a mixture of lignin model compounds in a relatively small number of generations could be related to its relatively high mutation rate. Across all adaptation experiments, we observed an average of 32 mutations after ~ 270–420 generations (~ 4 generations per subculture) (Table 2-S1), which is at least an order of magnitude higher than rates of occurrence observed in *Escherichia coli* under stress conditions [113]. Even in the high glucose condition, the average mutation occurrence was 19 mutations after ~ 290 generations in *R. opacus*. This high occurrence of mutations was not the result of any mutations in the DNA recombination and repair system, which are often found in hypermutator strains [114-116]. Notably, a large portion of SNPs observed in *R. opacus* were synonymous SNPs (Table 2-S2). While synonymous SNPs could potentially affect gene expression rates, protein folding, and thus protein functions [117], most synonymous SNPs in both the glucose and aromatic adaptation conditions were located in the same region of Plasmid 1, suggesting that these mutations are not related specifically to aromatic tolerance and utilization.

The most frequently mutated genes in *R. opacus* were related to redox reactions (Figure 2-3B). One such example was SOD, a highly conserved enzyme that protects cells from superoxide radicals (O_2^-) by converting them into molecular oxygen (O_2) or hydrogen peroxide (H_2O_2) [118]. We also observed multiple mutations affecting cytochrome ubiquinol oxidase in all phenolic compound adapted strains, but not in the high glucose or BEN adapted strains. This gene was also

mutated in a previous *R. opacus* adaptation experiment using phenol as a sole carbon source [86]. Cytochrome ubiquinol oxidase catalyzes the terminal step in one of the electron transport chains present in *Rhodococcus* strains, which reduces O_2 to H_2O [119]. The exact mechanistic contribution of these two mutated enzymes to the evolved aromatic phenotypes is not fully understood, but the known utilization of superoxide radicals and O_2 during aromatic consumption may provide some insights. Super-oxide radicals are generated from O_2 by aromatic degradation enzymes, including monooxygenase funneling enzymes (e.g., HBA mono-oxygenase) [102] and ring splitting dioxygenases (e.g., catechol 1,2-dioxygenase) [101], to facilitate aromatic catabolism. Our observation that PVHG6 had a 56% decreased SOD activity relative to the WT strain suggests a greater availability of superoxide radicals for aromatic degradation enzymes in PVHG6 than the WT strain.

By performing transcriptomic analyses on strains grown on single compounds, we were able to identify distinct funneling pathways for five lignin model compounds (Figure 2-5) and two aromatic by-product detoxification pathways (Figure 2-S19). The responses of funneling pathways to individual compounds and the mixture (Figure 2-5A), together with a distinct consumption pattern of lignin model compounds in the mixture (Figure 2-4B), suggest that *R. opacus* can differentially and specifically regulate aromatic utilization pathways in response to specific compounds. Carbon catabolite repression in *Rhodococcus* strains has been reported between aromatic compounds such as benzoate and phthalate [120] and 4-methoxybenzoate and 2-ethoxyphenol [121]. Relatedly, we observed differential upregulation of guaiacol degradation cluster #2 between the GUA growth condition and the mixture in both WT and PVHG6 strains (Figure 2-5A), suggesting that the presence of other aromatic compounds may delay guaiacol consumption (Figure 2-4B) by reducing the expression level of guaiacol degradation cluster #2.

Additionally, two pathways for a by-product of VAN and GUA catabolism, formaldehyde, in addition to the gluconeogenesis and pentose phosphate pathway genes that produce required enzymatic substrates, were all upregulated in both strains in the VAN, GUA, and mixture growth conditions (Figure 2-S19). Several transporter genes were also upregulated, and gene knockout experiments demonstrated their importance for aromatic utilization or tolerance, as suggested previously [86]. Together, these results show that *R. opacus* encodes a complex catabolic network for aromatic utilization, which includes a number of aromatic transporters and metabolic pathways that are regulated in a compound-specific manner.

In our previous report, we performed adaptive evolution using phenol as a sole carbon source [86]. Because lignin depolymerization generates heterogeneous aromatic mixtures, understanding tolerance and utilization mechanisms for a greater diversity of aromatic compounds and mixtures is necessary for lignin valorization. Accordingly, our current work tackles a substantially expanded set of adaptive evolution experiments on multiple lignin model compounds (including phenol) and their mixtures. From this large number of evolved trajectories, we chose six adaptive evolution experiments which represented both substrate diversity and mixture complexity and performed comparative multi-omic analyses of selected mutant strains that were grown in these diverse aromatic carbon sources. Our comprehensive, multi-omic analyses allowed for elucidation of improved aromatic catabolism in evolved *R. opacus* strains as well as identification of multiple compound-specific pathways, facilitating future lignin valorization efforts.

2.5 Materials and Methods

2.5.1 Chemicals and strains

Chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise indicated, and *Rhodococcus opacus* PD630 (DSMZ 44193) was used as the ancestral strain (WT) for comparison to all mutated strains. For all growth experiments (unless otherwise indicated), cells were grown at 30 °C and 250 rpm in one of two defined minimal salts media (A or B; medium composition as previously described [106]). Culture media were filter-sterilized using a 0.22 μm filter, and pH was adjusted to 7.0 using 6 M HCl or 2 M NaOH. Carbon and nitrogen sources were added as sterile-filtered stock solutions. *R. opacus* strains were maintained on tryptic soy broth (TSB) plates supplemented with 1.5% agar.

2.5.2 Adaptive evolution and growth assays of adapted strains

R. opacus was serially passaged on combinations of protocatechuate (PCA), guaiacol (GUA), phenol (PHE), 4-hydroxybenzoate (HBA), and vanillate (VAN) as carbon sources, and it was also passaged on each compound individually (Figure 2-1). *R. opacus* was also adapted on sodium benzoate (BEN) and high concentrations of glucose individually as sole carbon sources. Minimal medium A was used for adaptation experiments. To start adaptation experiments, a single colony of the ancestral strain from a TSB agar plate was grown at 30 °C in 2 mL of minimal medium A supplemented with 0.3 g/L PHE for 24–48 h, followed by a subculture using an initial OD_{600} of 0.05 in 10 mL of the same media for 24 h. Next, cells were centrifuged at 3000 g for 10 min and resuspended in minimal medium A without carbon or nitrogen sources, and adaptation experiments were performed at a 10 mL culture volume in a 50 mL glass culture tube with the corresponding carbon source(s) (Table 2-S1) and 1 g/L ammonium sulfate as the nitrogen source. For passaging, cultures were started at an initial cell density (OD_{600}) of 0.03. After ~ 30 passages, the initial cell density was increased to 0.05 for all adaptation experiments. Cells were passaged after reaching mid-exponential phase (OD_{600}> 0.3), and cells from each adaptation experiment at

each passage were stored at −80 °C in 15% glycerol. At each passage, the total aromatic concentration was increased by 1–5% until we did not observe growth after three days. If cells did not grow after three days, the culture was restarted from the previous −80 °C frozen stock, and the overall aromatic concentration was decreased until cell growth could be recovered. Once growth was recovered, the concentration was held constant for multiple passages before increasing concentrations in following passages. Adaptation experiments were stopped when we observed no growth after restarting adaptations from frozen stocks or adaptations reached > 100 passages. See Table 2-S1for details on adaptation experiments.

To isolate colonies, adapted cultures were streaked on minimal medium plates supplemented with 1.5% agar, 1 g/L ammonium sulfate, and respective carbon sources from the adaptation experiment (concentrations as indicated in Table 2-S13). Six colonies were picked and grown at 30 °C in a 5 mL culture volume in a 50 mL glass tube with minimal medium A supplemented with 5 g/L glucose and 1 g/L ammonium sulfate for 36 h before storage at −80 °C in 15% glycerol. To compare the phenotype of the isolated colonies to that of the WT strain using their adaptation carbon sources (Figure 2-2), frozen stocks of each isolated strain were streaked on TSB plates and a single colony was grown in 2 mL of minimal medium A supplemented with lignin model compounds from their adaptation condition (concentrations as indicated in Table 2-S13) for 24–48 h, and then subcultured using an initial OD_{600} of 0.05 in 10 mL of lignin model compounds from their adaptation condition (concentrations as indicated in Table 2-S14) for 24 h. Next, cells were centrifuged at 3000 g for 10 min and resuspended in minimal medium A without carbon or nitrogen sources, and experiments were performed at 30 °C with a 10 mL culture volume in a 50 mL glass culture tube with adaptation carbon source(s) (concentrations as indicated in Figure 2-2) and 1 g/L ammonium sulfate as the nitrogen source. Growth was monitored using

absorbance at 600 nm (Abs_{600}). Absorbance measurements were made with black 96 well plates (Greiner BioOne) using a TECAN Infinite M200 Pro plate reader. OD_{600} values were obtained by converting the Abs_{600} of 200µL of cell culture to OD_{600} using an experimentally determined correlation curve (OD_{600}=1.975 * [Abs_{600}−0.04], R^2= 1.00). Specific growth rates were calculated by fitting at least four converted OD_{600} values in the log phase.

2.5.3 R. opacus cultures for targeted metabolomics and transcriptomics

For phenotyping of adapted strains outside of adaptation conditions (Figures 2-4A, C), targeted metabolomics experiments (Figure 2-4B), and transcriptomic experiments (Figures 2-5, 2-6), cultures were started as follows. −80 °C frozen stocks of isolated colonies were streaked onto TSB agar plates and grown at 30 °C for three days. Using a loopful of cells from a TSB plate, seed cultures were grown at 30 °C in 2 mL of minimal medium A or B supplemented with lignin model compounds (0.2 g/L of BEN, GUA, PHE, HBA, and VAN, 1 g/L total aromatics) as carbon sources for 24 h, and then subcultured in 10 mL of the same media for 24 h with an initial cell density (OD_{600}) of 0.05. Next, cells were centrifuged at 3000 g for 10 min and resuspended in minimal medium A or B without carbon or nitrogen sources, and experiments were performed at 30 °C with a 10 mL culture volume in a 50 mL glass culture tube with respective carbon source(s) (see below for concentrations) and 1 g/L ammonium sulfate as the nitrogen source. For phenotyping of strains outside of adaptation conditions, 0.5 g/L of BEN, GUA, PHE, HBA, and VAN (2.5 g/L total aromatics), 0.6 g/L of BEN, GUA, PHE, HBA, and VAN (3.0 g/L total aromatics), or a combination of aromatics and glucose were used as carbon sources (at concentrations as indicated in figure captions). For targeted metabolomics experiments, 0.5 g/L of BEN, GUA, PHE, HBA, and VAN (2.5 g/L total aromatics) were used as carbon sources. For growth assays using individual compounds as sole carbon sources, six different concentrations of BEN (5, 6, 7, 8, 9, and 10 g/L),

GUA (1.25, 1.5, 1.75, 2, 2.25, and 2.5 g/L), PHE (0.75, 1, 1.25, 1.5, 1.75, and 2 g/L), HBA (2, 2.25, 2.5, 2.75, 3, and 3.25 g/L), and VAN (1.25, 1.5, 1.75, 2, 2.25, and 2.5 g/L) were used. For transcriptomic experiments, seven different growth media were used: 0.5 g/L of BEN, 0.5 g/L of GUA, 0.5 g/L of PHE, 0.5 g/L of HBA, 0.5 g/L of VAN, and 1 g/L of glucose; and 0.5 g/L of BEN, GUA, PHE, HBA, and VAN (2.5 g/L total aromatics). Carbon sources are listed in the respective figure or table. Growth was monitored using Abs_{600} as described previously.

2.5.4 DNA extraction and sequencing library preparation

DNA was extracted using a standard bead-beating, phenol-chloroform extraction methodology [86], and DNA concentrations were quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific). The DNA was prepared for sequencing using a Nextera Tagmentation Kit (Illumina) [122, 123]. Samples were barcoded using a KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems). The resulting PCR products were cleaned using Agencourt AMPure XP Beads (Agencourt Bioscience Corporation) and stored at −20 °C until use. For sequencing, samples were pooled together in equimolar ratios and diluted to a final DNA concentration of 10 nM using nuclease free water.

2.5.5 Genome sequencing and SNP analysis

For whole genome sequencing (Figure 2-3), 20μL of the pooled barcoded DNA samples were submitted for sequencing at the Center for Genome Sciences & Systems Biology, Washington University in St. Louis School of Medicine. Samples were paired-end sequenced using an Illumina Hi-Seq 2500 System. Raw Illumina paired-end reads were demultiplexed using barcodes and quality trimmed using Trimmomatic [124]. Trimmed reads were mapped to the Chinese Academy of Science reference genome of *R. opacus* (ASM59954v1) [95] with the

updated RefSeq annotation (GCF_000599545.1) using Bowtie2 [125]. The resulting SAM files were converted to BAM files and indexed using SAMtools [126]. SNPs and INDELs were called using Pilon [127], and SNPs also appearing in the WT strain were removed using an in-house Python script before quality filtering using GATK [128]. Trimmed reads were also *de novo* assembled using SPAdes [129], and the resulting genome assembly was used to confirm the presence or absence of SNPs using BLAST+ [130]. Additionally, *de novo* confirmed SNPs were compared to another reference genome from the Broad Institute (ASM23433v1) [131] to filter out SNPs that were identical to the Broad Institute reference genome. Gene Ontologies [98] were inferred with Blast2GO [132] by finding the top 20 most similar protein sequences in the NCBI and UniProt Knowledgebase databases for every protein in the *R. opacus* genome using BLAST+ and InterProScan [133].

2.5.6 RNA extraction and rRNA depletion

RNA was extracted using a ZR Fungal/Bacterial RNA MiniPrep Kit (Zymo Research) and treated with TURBO DNase I (Ambion) for 30 min at 37 °C to remove DNA contamination. DNase-treated RNA was then cleaned using an RNA Clean & Concentrator Kit (Zymo Research). Samples were checked for DNA contamination by PCR amplification using primers specific to intergenic regions of the genome. Samples that were positive for DNA contamination were re-treated with TURBO DNase I and cleaned until no DNA was detected by PCR amplification. Next, RNA was quantified using a NanoVue Plus™ Spectrophotometer (Biochrom), and ribosomal RNA (rRNA) was depleted from total RNA using a Bacterial Ribo-Zero rRNA Removal Kit (Illumina). The resulting mRNA was converted to cDNA and barcoded using a reverse transcription protocol described previously [86]. cDNA samples were pooled together in equimolar ratios and diluted to a final cDNA concentration of 10 nM using nuclease free water.

2.5.7 Transcriptomic analysis

For transcriptomic analysis of WT and PVHG6 strains (Figures 2-5, 2-6), 20 μL of the pooled barcoded cDNA samples were submitted for sequencing at the Center for Genome Sciences & Systems Biology, Washington University in St. Louis School of Medicine. Samples were single-end sequenced using an Illumina Hi-Seq 2500 System. Raw single-end reads were demultiplexed, trimmed, and mapped in the same manner as the SNP analysis described previously. After indexing mapped reads, expression counts were calculated for each gene using featureCounts [134], and downstream differential gene expression analysis was performed using DESeq2 [135].

2.5.8 Aromatic consumption profiling

For cultures using a single carbon source, the concentration of each lignin model compound was estimated by comparing UV absorbance values to a standard curve for each lignin model compound (Figures 2-4C, 2-S22). 200μL of each cell culture was centrifuged at 16,000 g for 5 min, and the UV absorbance value of a 100-fold dilution of culture supernatant was measured in a UV transparent 96 well plate (ThermoScientific) using a TECAN Infinite M200 Pro plate reader, or in a quartz cuvette using a Nanodrop 2000 (Thermo Scientific). For cultures using the aromatic mixture (Figure 2-4B), aromatics in the culture supernatant were derivatized using methyl chloroformate [136] and quantified using a custom GC-MS-FID system (gas chromatography-mass spectrometry-flame ionization detector). Briefly, the culture supernatant was centrifuged at 16,000 g for 5 min, and 200 μL of the culture supernatant was mixed with 40 μL of 5.0% (w/w) sodium hydroxide solution, 200 μL of methanol, and 50 μL of pyridine. Methylchloroformate (50 μL) was added to the mixture in two 25 μL aliquots. Next, 400 μL of chloroform containing a decane internal standard was added to the sample, followed by the addition of 400 μL of 50 mM sodium bicarbonate solution to induce phase separation of the aqueous and organic layers. Samples

were vortexed between each step to ensure complete mixing. After phase separation, the organic phase was transferred to a GC vial with a 350 µL glass insert (Agilent), and samples were analyzed using an Agilent 7890A GC coupled to both an Agilent 5975C mass spectrometer containing a triple-axis detector and an Agilent G3461A FID detector with a methanizer (Activated Research Company). The Agilent 7890A GC was equipped with a Restek fused silica RTX-50 capillary column (30 m by 0.25 mm, 0.5 µm film thickness), and helium was used as the carrier gas. 1 µL of the organic phase was injected with a splitting ratio of 10:1 using the autosampler. For GC runs, the inlet was maintained at 250 °C, and the oven was held for 2 min at 40 °C, heated to 300 °C using a 5 °C/min ramp, and held at 300 °C for 5 min. All data was exported and analyzed using the Agilent ChemStation Software, and peak intensities were normalized to the decane internal standard. Lignin model compounds were identified based on retention time and mass spectral database searches using the Palisade Complete Mass Spectral Database (600 K edition, Palisade Mass Spectrometry, Ithaca, NY), and concentrations were determined using an external standard curve for each lignin model compound.

2.5.9 Lipid assay

For lipid production assays (Figure 2-7), growth experiments were performed in the same way as metabolomic and transcriptomic experiments with the following modification: the culture volume was 50 mL in a 250 mL baffled flask, the carbon sources for the experiment were 0.5 g/L of BEN, GUA, PHE, HBA, and VAN (2.5 g/L total aromatics), and the nitrogen source for the experiment was 0.125 g/L ammonium sulfate. Lipid was measured gravimetrically using a modified Folch lipid extraction [137]. To measure biomass content, 10 mL of cell culture was centrifuged at 4600 g for 15 min, washed twice with distilled water, dried at 60 °C for two days, and weighed to determine the dry cell weight. For lipid extractions, 30 mL of cell culture was

centrifuged at 4600 g for 15 min, washed twice with distilled water, and resuspended in 4 mL of a chloroform/methanol solution (2:1). Next, the solution was sonicated at 20% amplitude using the program: 1 s on and 1 s off for 5 min. The solution was partitioned into aqueous and organic phases using 0.2 volumes of 0.9% NaCl solution. The organic phase was separated from the aqueous phase, dried under vacuum, and weighed to determine the total lipid amount.

2.5.10 Generation of R. opacus knockout mutants

Knockout mutants were generated via homologous recombination in the wild type *R. opacus* strain as described previously [138]. Briefly, a helper plasmid expressing a pair of recombinases (pDD120) was introduced via electroporation into WT *R. opacus*. Electrocompetent cells were made from this strain as previously described [106]. An integration vector containing an antibiotic resistance marker and DNA regions homologous to each target gene (~ 500 bp per homologous arm) was then introduced into the strain harboring the recombinase helper plasmid by electroporation [138]. Successful recombination was confirmed by colony PCR. See Table 2-S15 for plasmids utilized for knockouts and Table 2-S16 for all strains generated. Plasmid DNA was isolated from *Escherichia coli* DH10B for electroporation into *R. opacus* using the Zyppy Plasmid Miniprep Kit (Zymo).

2.5.10 Data Availability

Genomic and transcriptomic sequences have been deposited with the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under accession number SRP131196.

2.6 Supplemental Figures

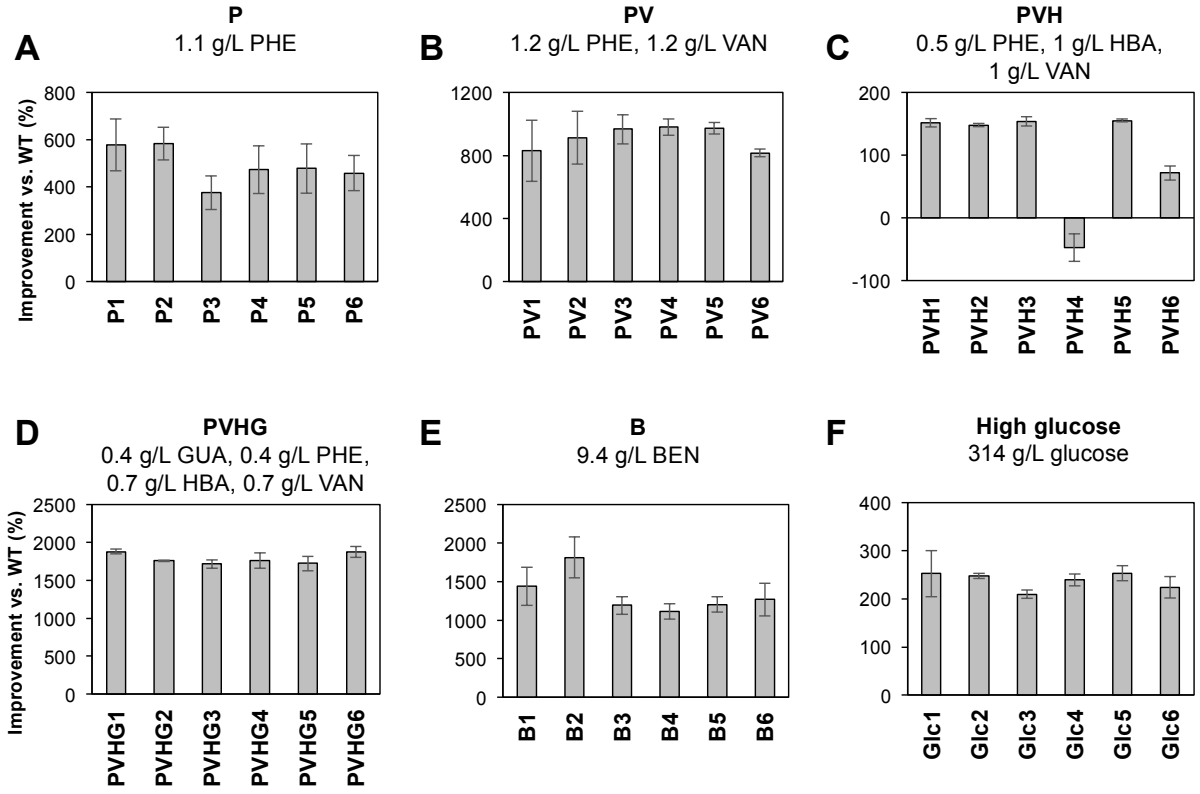


Figure 2-S1: Cell density comparison between WT and adapted strains. Values are based on the data in Figure 2-2 and reported as improvement of final cell density (OD_{600}) in the adapted strain compared to OD_{600} of the WT strain. % improvement = ($Adapted_{OD600}$ – WT_{OD600}) / WT_{OD600} x 100%, where $Adapted_{OD600}$ and WT_{OD600} is the OD_{600} of the adapted and WT strains, respectively, after 43 hours of growth. Cells were grown in a minimal salts medium supplemented with the listed carbon sources as the sole carbon sources and 1 g/L ammonium sulfate as the nitrogen source. (A) The phenol adapted strains (P1-6). (B) The phenol and vanillate adapted strains (PV1-6). (C) The phenol, vanillate, and 4-hydroxybenzoate adapted strains (PVH1-6). (D) The phenol, vanillate, 4-hydroxybenzoate, and guaiacol adapted strains (PVHG1-6). (E) The sodium benzoate adapted strains (B1-6). (F) The high glucose adapted strains (Glc1-6). Bars represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

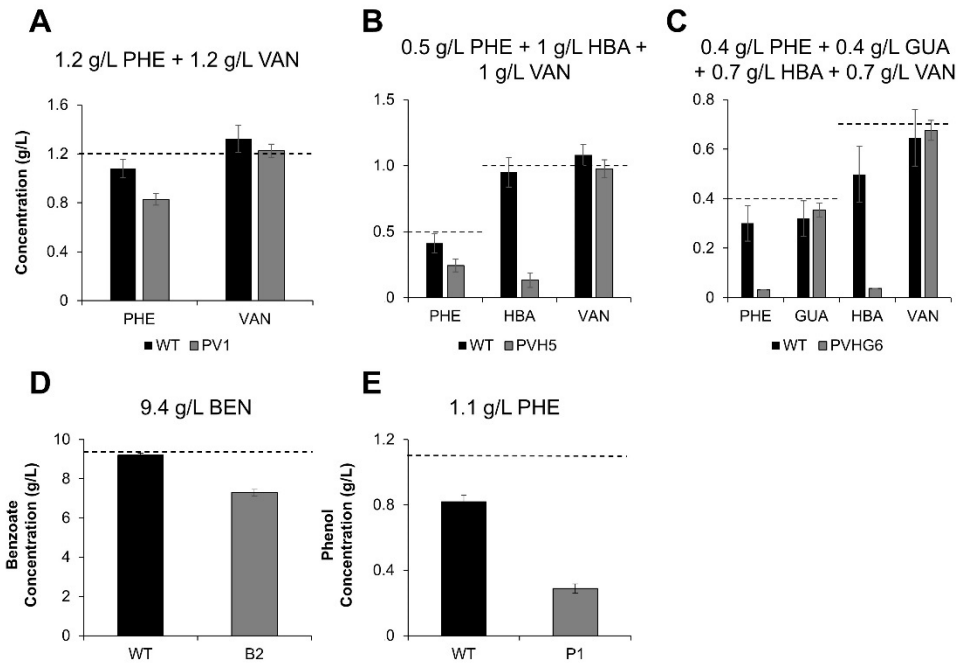


Figure 2-S2: Aromatic concentrations in culture supernatants after 43 hours for WT and aromatic compound adapted strains. A. The PHE and VAN adapted strain PV1 compared to the WT strain using 1.2 g/L PHE and 1.2 g/L VAN. B. The PHE, VAN, and HBA adapted strain PVH5 compared to the WT strain using 0.5 g/L PHE, 1 g/L HBA, and 1 g/L VAN. C. The PHE, VAN, HBA, and GUA adapted strain PVHG6 compared to the WT strain using 0.4 g/L GUA, 0.4 g/L PHE, 0.7 g/L HBA, and 0.7 g/L VAN. D. The sodium benzoate adapted strain B2 compared to the WT strain using 9.4 g/L BEN. E. The phenol adapted strain P1 compared to the WT strain using 1.1 g/L PHE. Aromatic concentrations were measured by derivatizing culture supernatants using methyl chloroformate followed by GC-MS-FID (see Materials and Methods for details). Cells were grown in a minimal medium A supplemented with the listed carbon sources (same as Fig. 2) as the sole carbon sources and 1 g/L ammonium sulfate as the nitrogen source. Bars represent the average of three biological replicates and error bars represent one standard deviation. The dotted lines indicate the initial aromatic concentration. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

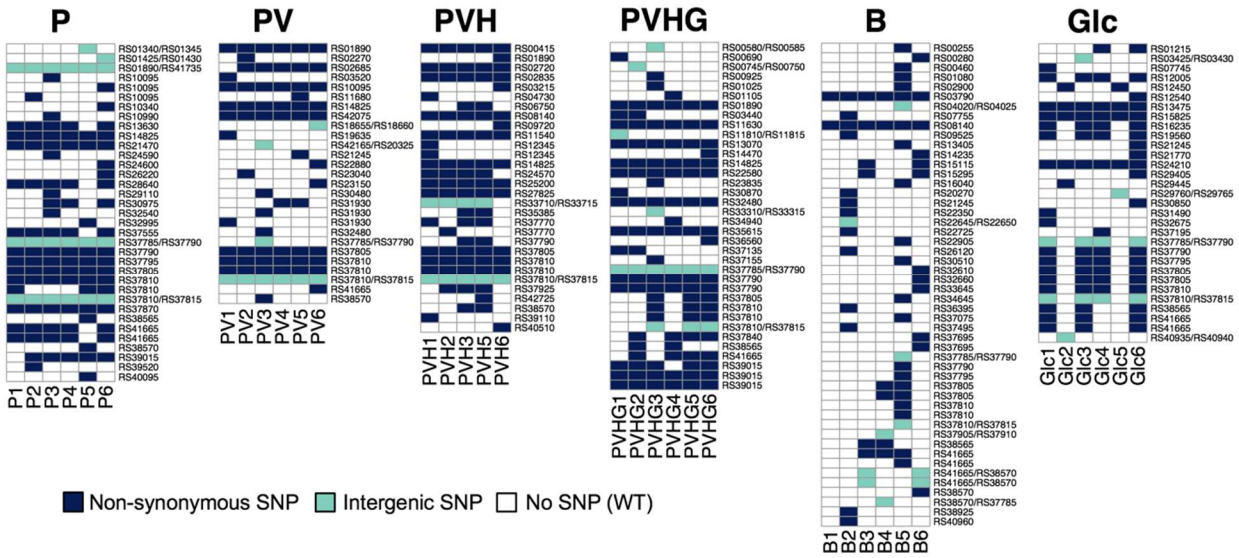


Figure 2-S3: Genes with non-synonymous and intergenic SNPs in adapted strains separated by adaptation conditions. Some genes have multiple mutations, so each row represents a specific mutation. Strain names indicate the adaptation condition from which the strain originates: P = phenol; PV = phenol and vanillate; PVH = phenol, vanillate, and 4-hydroxybenzoate; PVHG = phenol, vanillate, 4-hydroxybenzoate, and guaiacol; B = sodium benzoate; Glc = glucose.

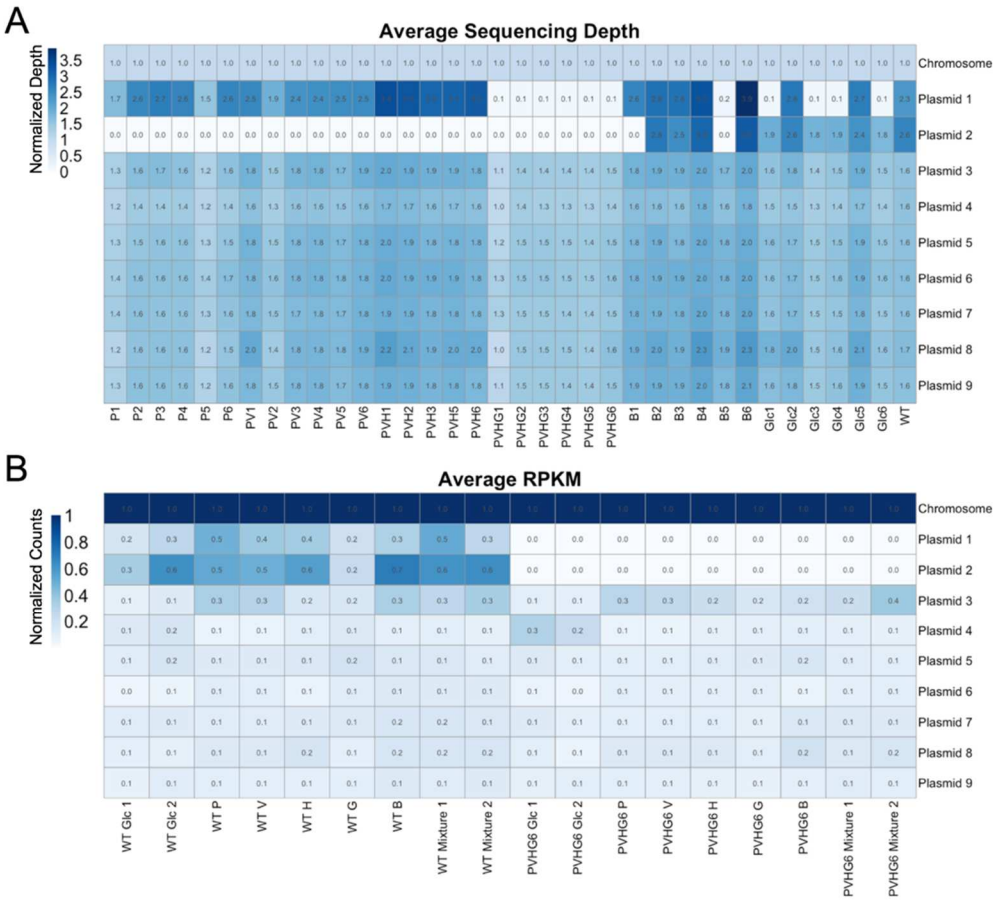


Figure 2-S4: Average sequencing depth for all sequenced strains (35 mutants and WT) and normalized counts in all transcriptomic samples (PVHG6 and WT). A. The average sequencing depth per base pair in the chromosome or plasmids. Average sequencing depth values were calculated by summing the total sequencing depth of all the individual base pairs on the chromosome or plasmid and dividing by the number of base pairs on the respective chromosome or plasmid. Sequencing depths were then normalized by the average depth of the chromosome for each strain. B. Average normalized counts were calculated by summing the total number of reads per kilobase of transcript per million mapped reads (RPKM) for each gene on the chromosome or plasmid and dividing by the number of genes on the respective chromosome or plasmid. The average RPKM values were then normalized by the average RPKM value of the chromosome for each strain. Glc = 1 g/L glucose, P = 0.5 g/L phenol, V = 0.5 g/L vanillate, H = 0.5 g/L 4-hydroxbenzoate, G = 0.5 g/L guaiacol, B = 0.5 g/L sodium benzoate, and Mixture = 0.5 g/L of phenol, vanillate, 4-hydroxybenzoate, guaiacol, and sodium benzoate (2.5 g/L total aromatics). For Glc and Mixture, "1" represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Mixture), while "2" represents cells harvested at mid-exponential phase (13 hours for Glc; 32 hours for Mixture). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT P, 21 hours for PVHG6 P, 19 hours for WT and PVHG6 G, 11 hours for WT and PVHG6 H, 12 hours for WT and PVHG6 B, and 24 hours for WT and PVHG6 V). All transcriptomics experiments were performed in minimal medium A. For B, numerical values represent the average of three biological replicates.

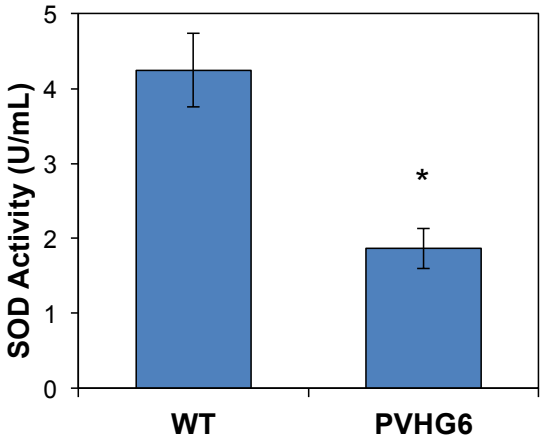


Figure 2-S5: Comparison of superoxide dismutase (SOD) activity from WT and PVHG6 cell lysates. Measured relative SOD activity from WT and PVHG6 cell lysates. Asterisk indicates significant difference from WT SOD activity (P = 0.002, one mean, two-tailed Student's t-test). Cells were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured in 10 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours to an OD_{600} of ~0.5. Cells were grown in 100 mL of minimal medium A supplemented with 1 g/L glucose and 1 g/L ammonium sulfate to mid-exponential phase (OD_{600} = 0.44-0.49 for both strains; three replicates). Cultures (the same number of cells equivalent to OD_{600} = 0.44 for both strains by adjusting volumes for harvesting) were centrifuged at 1000g and 4 °C for 15 minutes, resuspended in 300 µL of EMD Millipore BugBuster Protein Extraction Reagent, and incubated at room temperature for 30 minutes. Next, the lysed cells were centrifuged at 16,000g and 4 °C for 10 minutes, and a ten-fold dilution of supernatant was used for testing. SOD activity was measured using the Invitrogen Superoxide Dismutase Colorimetric Activity Kit. SOD activity was compared to a standard curve generated using bovine erythrocyte superoxide dismutase in the Kit. Bars represent the average of three biological replicates and error bars represent one standard deviation.

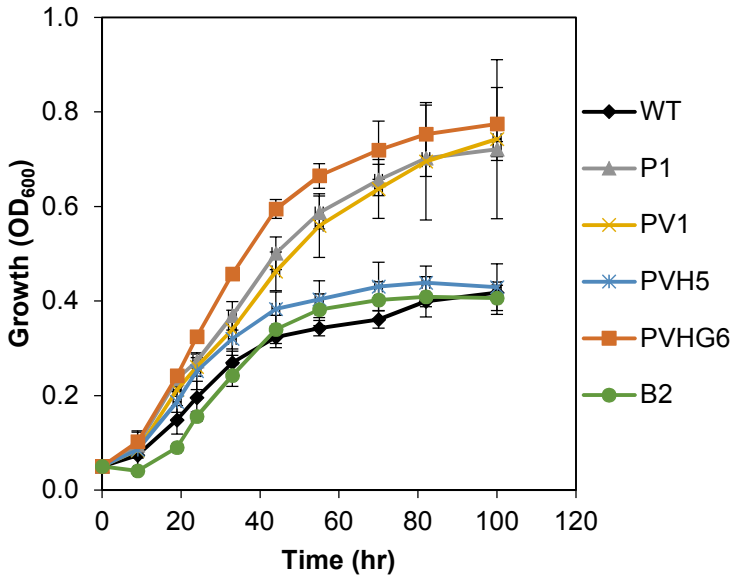


Figure 2-S6: Growth of WT and adapted strains in a mixture of lignin model compounds. Growth of strains isolated from adapted cultures in 0.6 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (3 g/L total aromatics) using minimal medium A. Cells were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured in 10 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours to an OD_{600} of ~0.5. For all data shown, cultures were grown in 10 mL of minimal medium A supplemented with 1 g/L ammonium sulfate as the nitrogen source with an initial OD_{600} of 0.05. Points represent the average of three biological replicates and error bars represent one standard deviation.

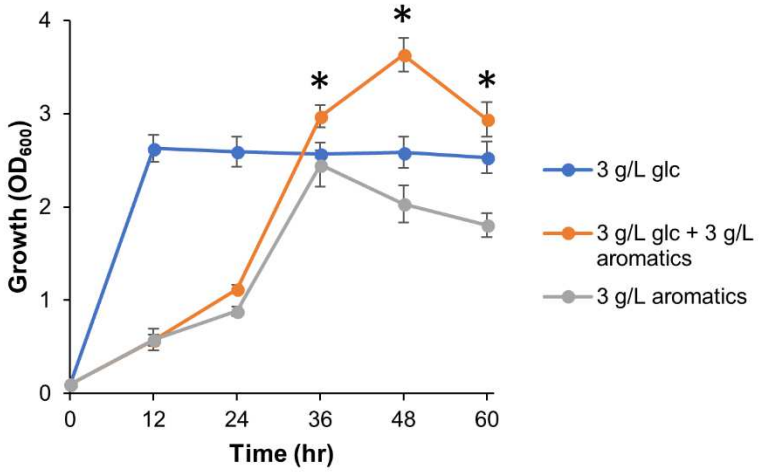


Figure 2-S7: Growth of PVHG6 using glucose, lignin model compounds, and a mixture of glucose and lignin model compounds as carbon sources. Growth of PVHG6 in 0.6 g/L each of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (3 g/L total aromatics), 3 g/L glucose (glc), or both glucose and aromatics as carbon sources. Cells were initially cultured in 2 mL of minimal medium B supplemented with 0.2 g/L each of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured in 10 mL of minimal medium B supplemented with 0.2 g/L each of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours to an OD_{600} of ~0.5. For all data shown, cultures were grown in 10 mL of minimal medium B supplemented with 1 g/L ammonium sulfate as the nitrogen source with an initial OD_{600} of 0.1. Points represent the average of three biological replicates and error bars represent one standard deviation. Asterisk (*) indicates a significantly higher OD_{600} value ($P < 0.05$, one mean, two-tailed Student's t-test) for the 3 g/L glucose + 3 g/L aromatics condition than both the 3 g/L glucose and 3 g/L aromatics conditions. Analysis of derivatized culture supernatants by GC-MS-FID at the 60 hour time point showed that all aromatics were consumed for both the 3 g/L aromatics and the 3 g/L glucose + 3 g/L aromatics conditions.

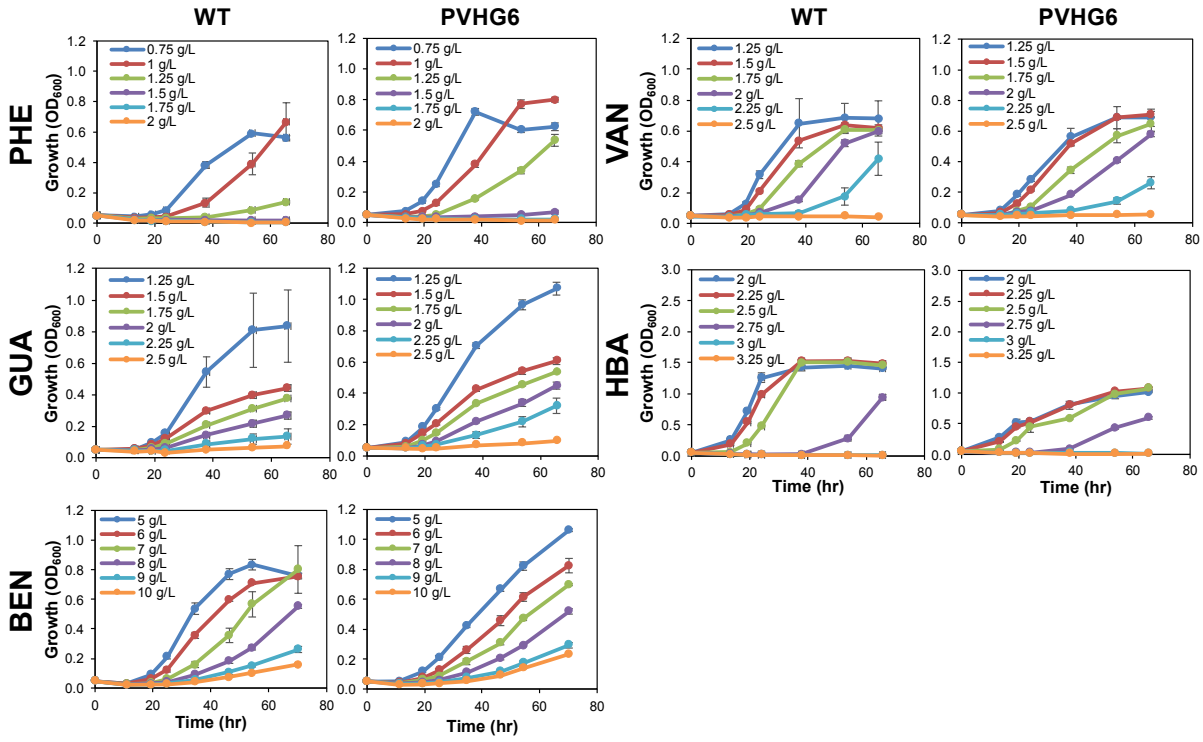


Figure 2-S8: Growth comparison of PVHG6 and WT strains using individual lignin model compounds. Growth comparison of PVHG6 and WT strains in different concentrations of individual lignin model compounds. The individual compound used for growth is listed next to the y-axis, and the strain is listed above each column. Cells were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured in 10 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours to an OD_{600} of ~0.5. For all data shown, cultures were grown in 10 mL of minimal medium A supplemented with 1 g/L ammonium sulfate as the nitrogen source with an initial OD_{600} of 0.05. Points represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, GUA = guaiacol, VAN = vanillate, HBA = 4-hydroxybenzoate, BEN = sodium benzoate.

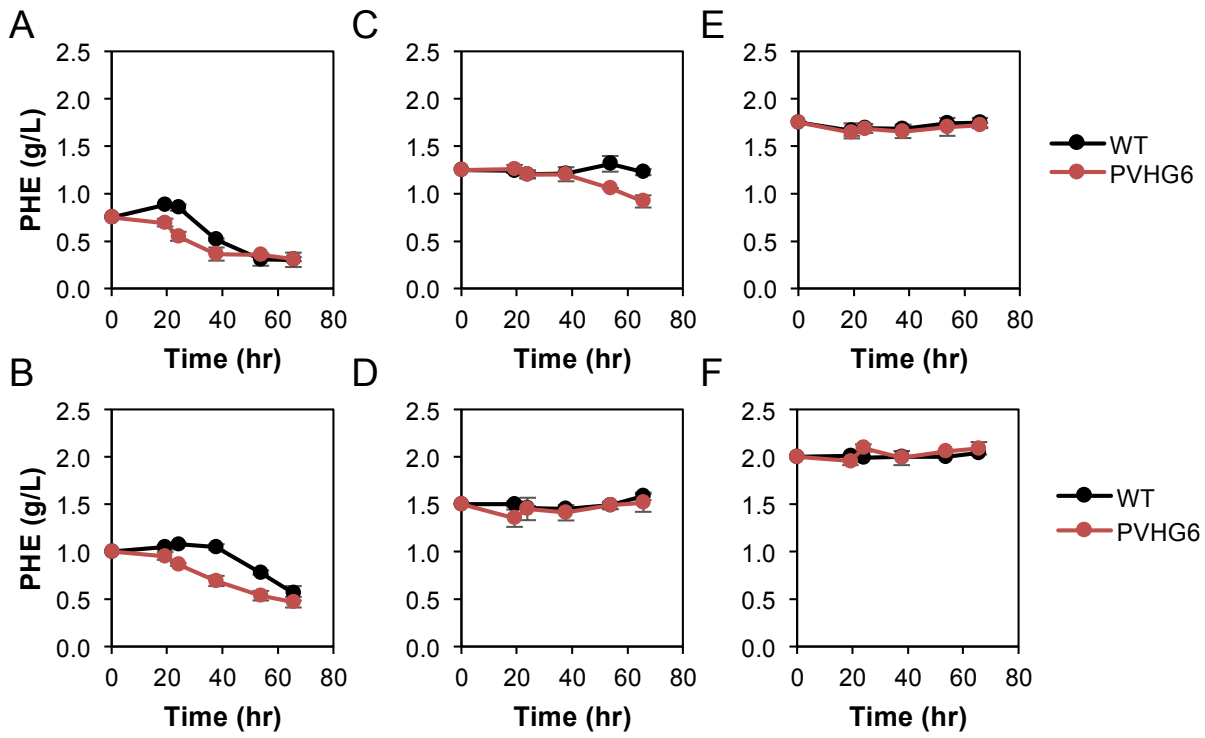


Figure 2-S9: Phenol (PHE) consumption profiles of WT and PVHG6 strains grown in different PHE concentrations. A. 0.75 g/L. B. 1 g/L. C. 1.25 g/L. D. 1.5 g/L. E. 1.75 g/L. F. 2 g/L. Points represent the average of three biological replicates and error bars represent one standard deviation. Phenol concentration was measured in culture supernatants using absorbance at 270 nm. See Fig. 2-S8 for growth conditions and profiles.

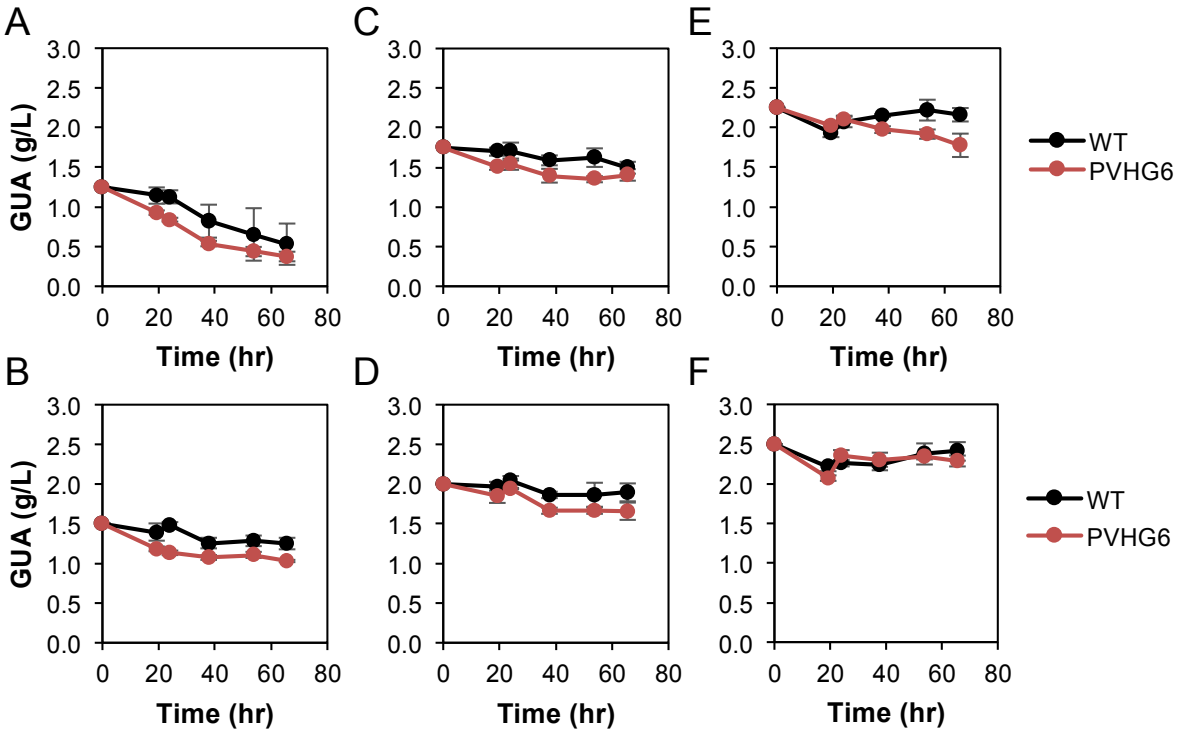


Figure 2-S10: Guaiacol (GUA) consumption profiles of WT and PVHG6 strains grown in different GUA concentrations. A. 1.25 g/L. B. 1.5 g/L. C. 1.75 g/L. D. 2 g/L. E. 2.25 g/L. F. 2.5 g/L. Points represent the average of three biological replicates and error bars represent one standard deviation. Guaiacol concentration was measured in culture supernatants using absorbance at 274 nm. See Fig. 2-S8 for growth conditions and profiles.

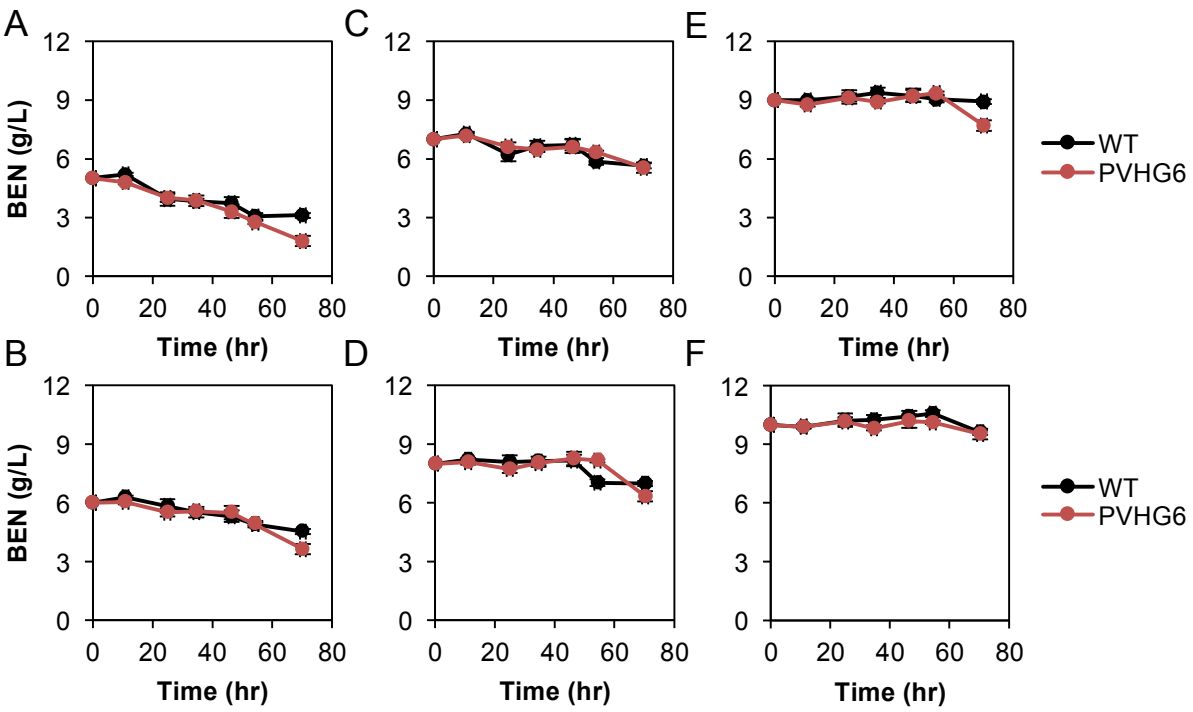


Figure 2-S11: Sodium benzoate (BEN) consumption profiles of WT and PVHG6 strains grown in different BEN concentrations. A. 5 g/L. B. 6 g/L. C. 7 g/L. D. 8 g/L. E. 9 g/L. F. 10 g/L BEN. Points represent the average of three biological replicates and error bars represent one standard deviation. Sodium benzoate concentration was measured in culture supernatants using absorbance at 225 nm. See Fig. 2-S8 for growth conditions and profiles.

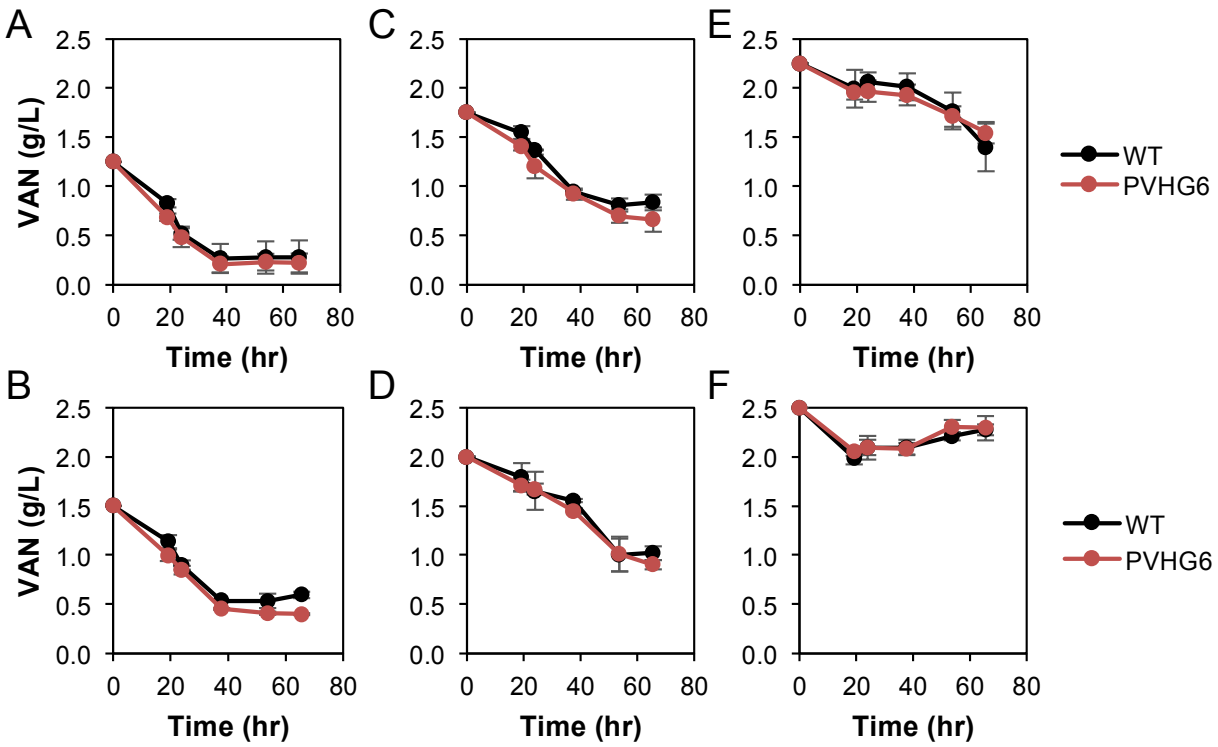


Figure 2-S12: Vanillate (VAN) consumption profiles of WT and PVHG6 strains grown in different VAN concentrations. A. 1.25 g/L. B. 1.5 g/L. C. 1.75 g/L. D. 2 g/L. E. 2.25 g/L. F. 2.5 g/L. Points represent the average of three biological replicates and error bars represent one standard deviation. Vanillate concentration was measured in culture supernatants using absorbance at 284 nm. See Fig. 2-S8 for growth conditions and profiles.

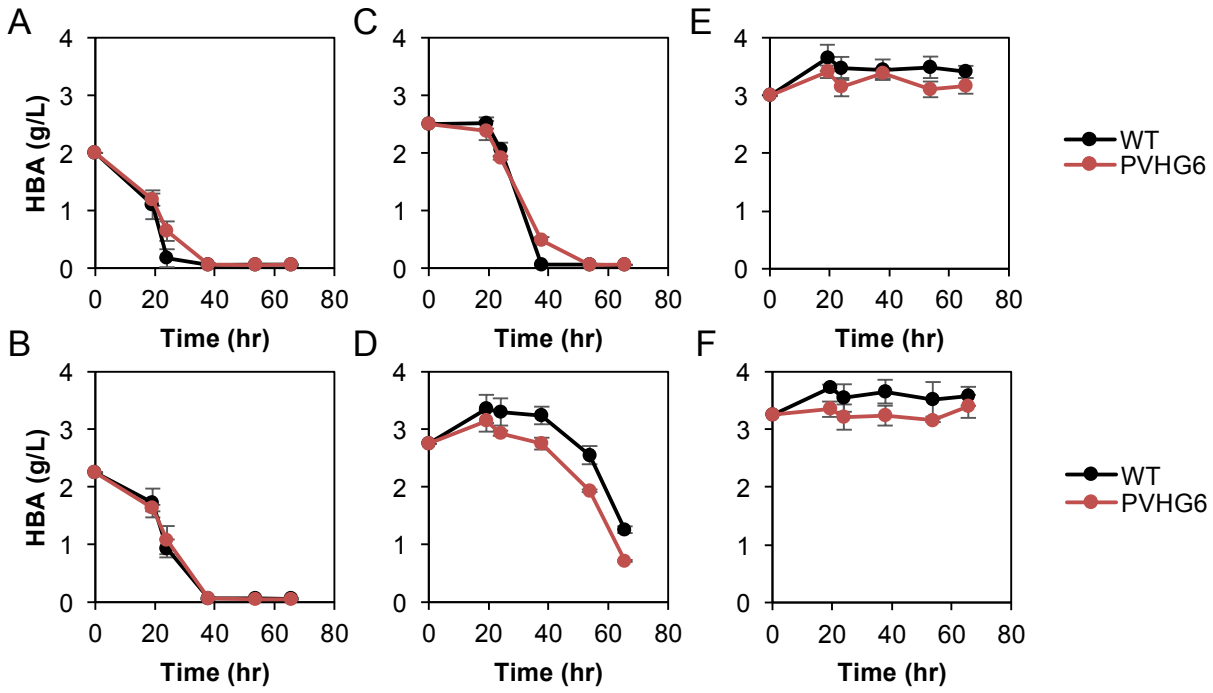


Figure 2-S13: 4-hydroxybenzoate (HBA) consumption profiles of WT and PVHG6 strains grown in different HBA concentrations. A. 2 g/L. B. 2.25 g/L. C. 2.5 g/L. D. 2.75 g/L. E. 3 g/L. F. 3.25 g/L. Points represent the average of three biological replicates and error bars represent one standard deviation. 4-hydroxybenzoate concentration was measured in culture supernatants using absorbance at 244 nm. See Fig. 2-S8 for growth conditions and profiles.

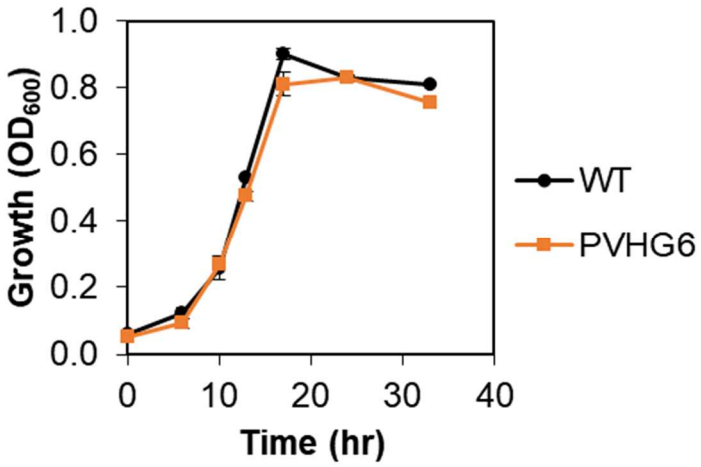


Figure 2-S14: Growth of WT and PVHG6 strains on 1 g/L glucose. Cells were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured into 10 mL of the same media for 24 hours to an OD_{600} of ~0.5. Cultures were grown in 10 mL of minimal medium A supplemented with 1 g/L glucose as the carbon source and 1 g/L ammonium sulfate as the nitrogen source with an initial OD_{600} of 0.05. Points represent the average of three biological replicates and error bars represent one standard deviation.

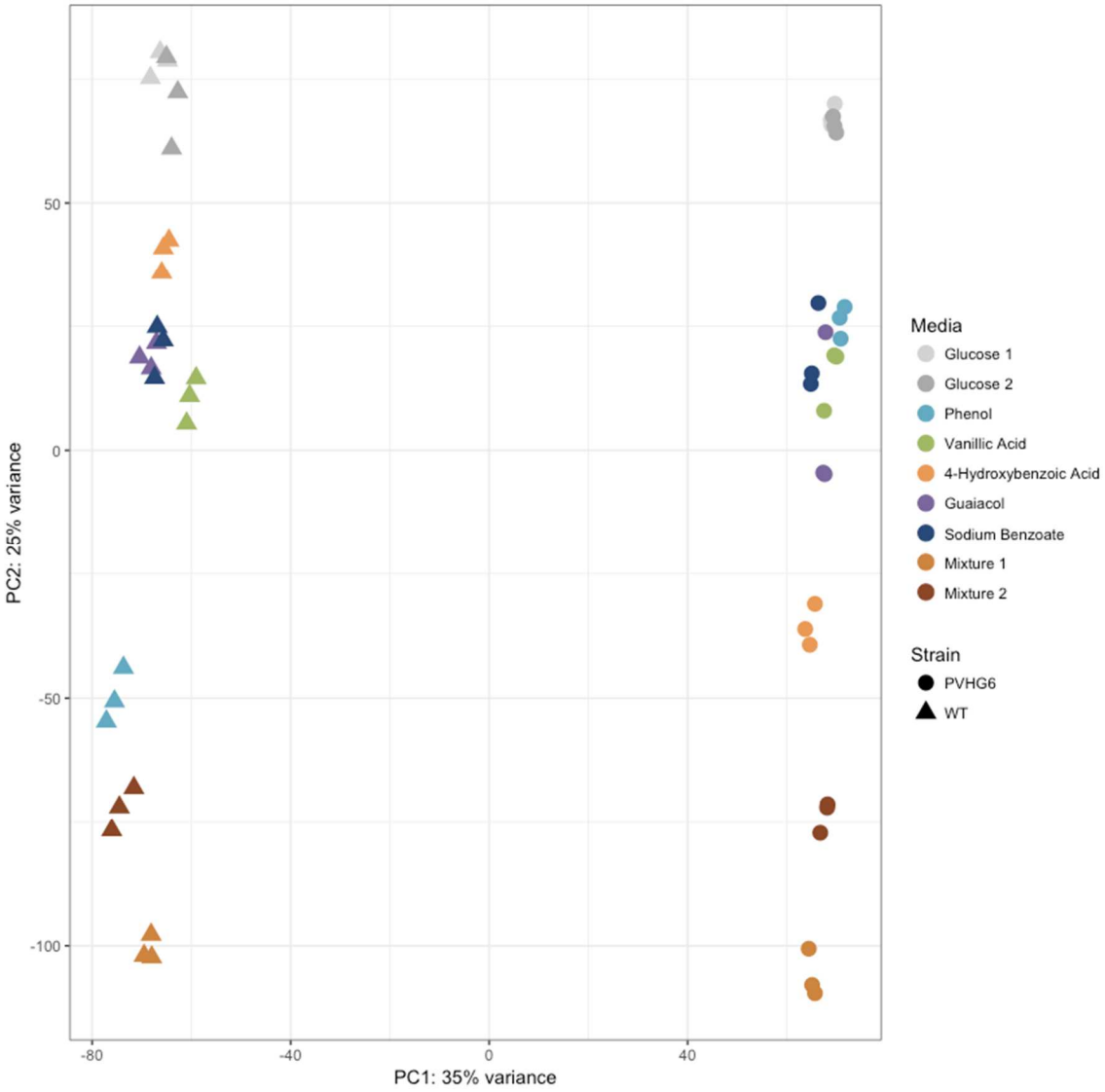


Figure 2-S15: Principal component analysis of all WT and PVHG6 transcriptomes. Principal component analysis was done using normalized counts of all genes in *R. opacus*. Glc = 1 g/L glucose, Multi = 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). Other compounds listed in legend were used at a concentration of 0.5 g/L. For Glc and Multi, t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc, 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE and 21 hours for PVHG6 PHE; 19 hours for WT and PVHG6 GUA; 11 hours for WT and PVHG6 HBA; 12 hours for WT and PVHG6 BEN; 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Each symbol represents a biological replicate. Analysis was performed with normalized counts using the variance stabilizing transformation from the DESeq2 software package. PC1 = 1st principal component and PC2 = 2nd principal component. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

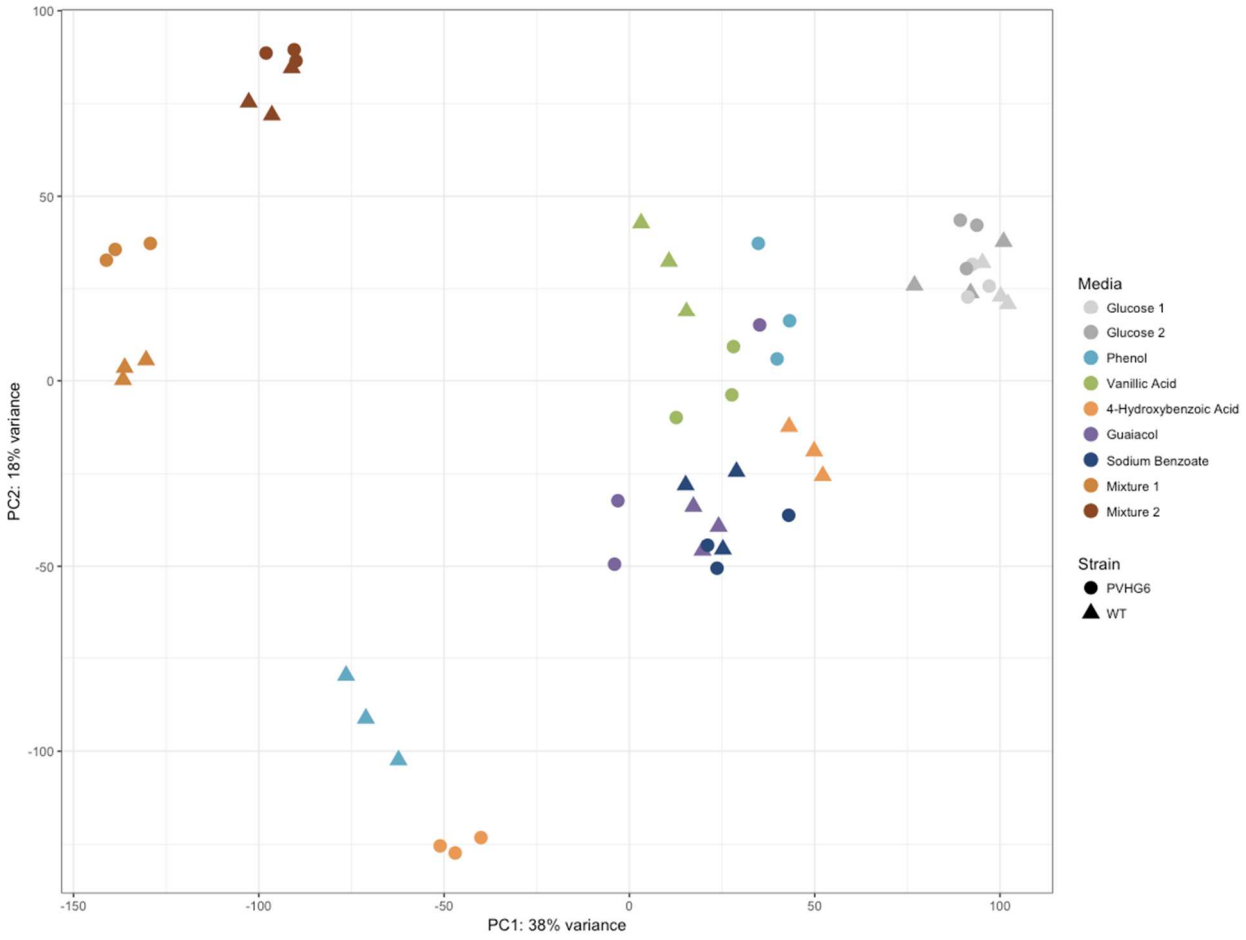


Figure 2-S16: Principal component analysis of all WT and PVHG6 transcriptomes with Plasmids 1 and 2 removed. Principal component analysis was done using normalized counts of all genes in *R. opacus* except those located on Plasmids 1 and 2. Glc = 1 g/L glucose, Multi = 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). Other compounds listed in legend were used at a concentration of 0.5 g/L. For Glc and Multi, t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc, 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE and 21 hours for PVHG6 PHE; 19 hours for WT and PVHG6 GUA; 11 hours for WT and PVHG6 HBA; 12 hours for WT and PVHG6 BEN; 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Each symbol represents a biological replicate. Analysis was performed with normalized counts using the variance stabilizing transformation from the DESeq2 software package. PC1 = 1st principal component and PC2 = 2nd principal component. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

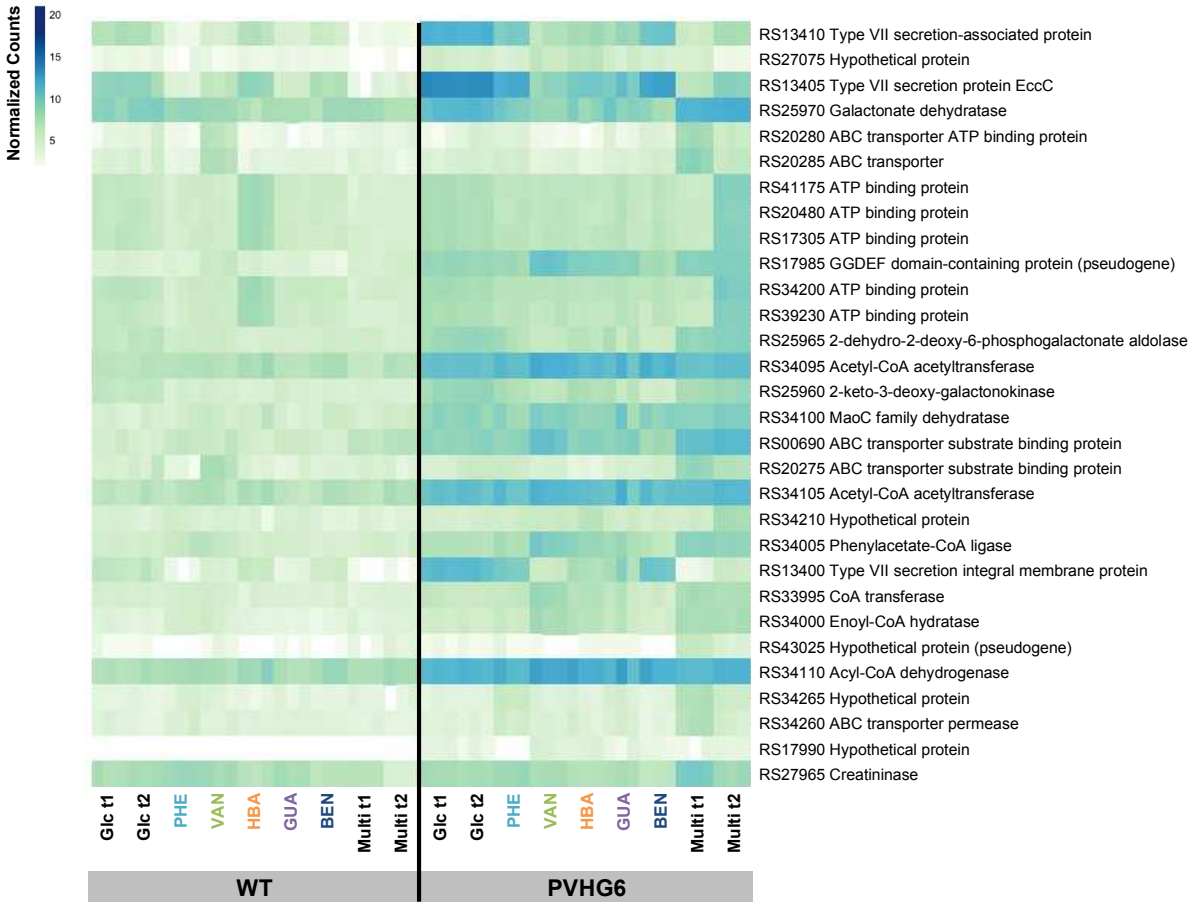


Figure 2-S17: Expression of the top 30 most upregulated genes in PVHG6 compared to the WT strain in the mixture of lignin model compounds. Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 g/L 4-hydroxbenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of phenol, vanillate, 4-hydroxybenzoate, guaiacol, and sodium benzoate (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc; 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE, 21 hours for PVHG6 PHE, 19 hours for WT and PVHG6 GUA, 11 hours for WT and PVHG6 HBA, 12 hours for WT and PVHG6 BEN, and 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Gene codes are from the NCBI reference sequence NZ_CP003949.1.

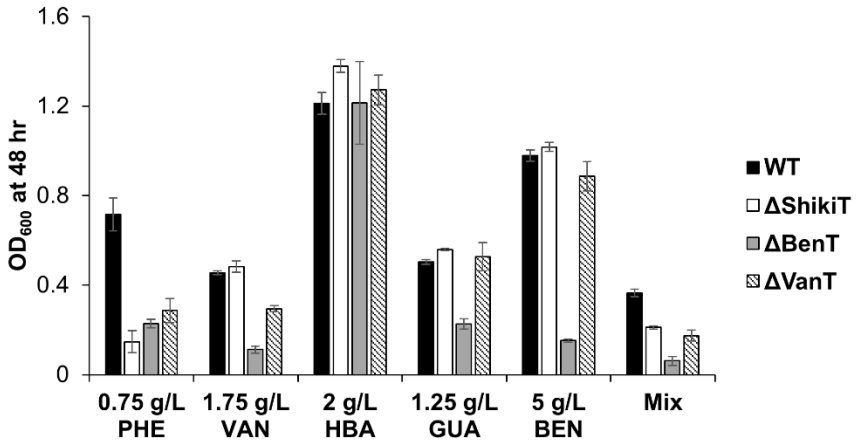


Figure 2-S18: Growth of transporter knockout mutants compared to that of the WT strain using aromatic carbon sources. Cells were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured into 10 mL of the same media for 24 hours to an OD_{600} of ~0.5. For all data shown, cultures were grown in 10 mL of minimal medium A supplemented with 1 g/L ammonium sulfate as the nitrogen source and the carbon source as indicated in the figure with an initial OD_{600} of 0.05. Bars represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, BEN = sodium benzoate, and Mix = 0.5 g/L of PHE, VAN, HBA, GUA and BEN (2.5 g/L total aromatics). ΔShikiT = RS31355 knockout mutant, ΔBenT = RS30810 knockout mutant, and ΔVanT = RS33590 knockout mutant.

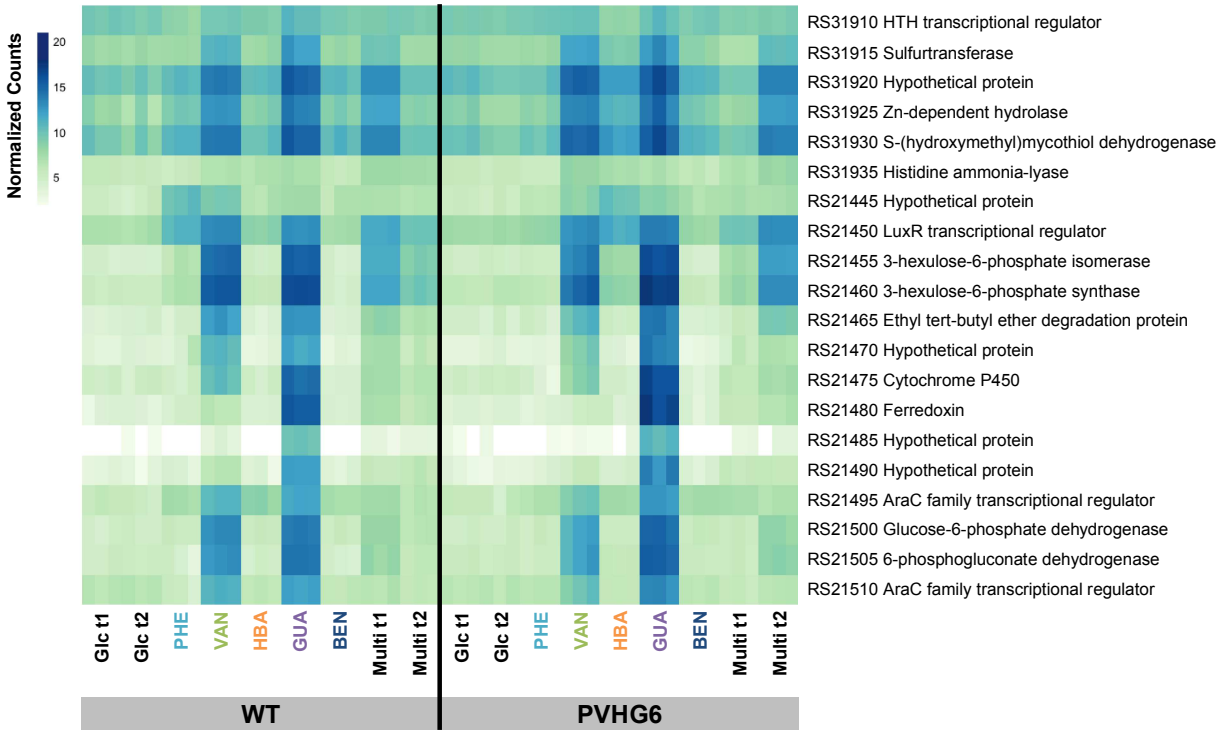


Figure 2-S19: Putative C1 metabolism during growth on vanillate and guaiacol. Heat map of normalized gene expression counts for genes putatively involved in C1 metabolism. Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 g/L 4-hydroxbenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of phenol, vanillate, 4-hydroxybenzoate, guaiacol, and sodium benzoate (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc; 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE, 21 hours for PVHG6 PHE, 19 hours for WT and PVHG6 GUA, 11 hours for WT and PVHG6 HBA, 12 hours for WT and PVHG6 BEN, and 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Gene codes are from the NCBI reference sequence NZ_CP003949.1.

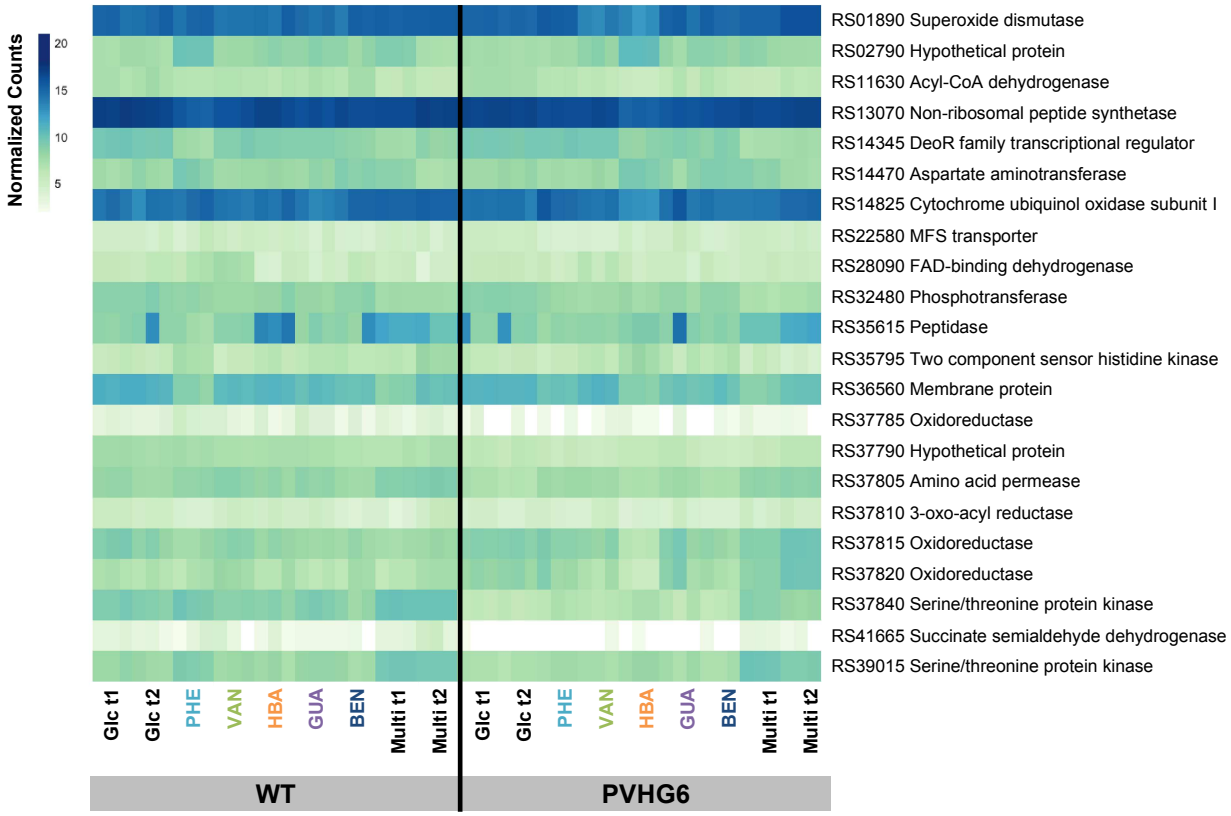


Figure 2-S20: Expression of genes with SNPs in PVHG6. Heat map of normalized gene expression counts for genes with SNPs in PVHG6. Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 g/L 4-hydroxbenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of phenol, vanillate, 4-hydroxybenzoate, guaiacol, and sodium benzoate (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc; 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE, 21 hours for PVHG6 PHE, 19 hours for WT and PVHG6 GUA, 11 hours for WT and PVHG6 HBA, 12 hours for WT and PVHG6 BEN, and 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Gene codes are from the NCBI reference sequence NZ_CP003949.1.

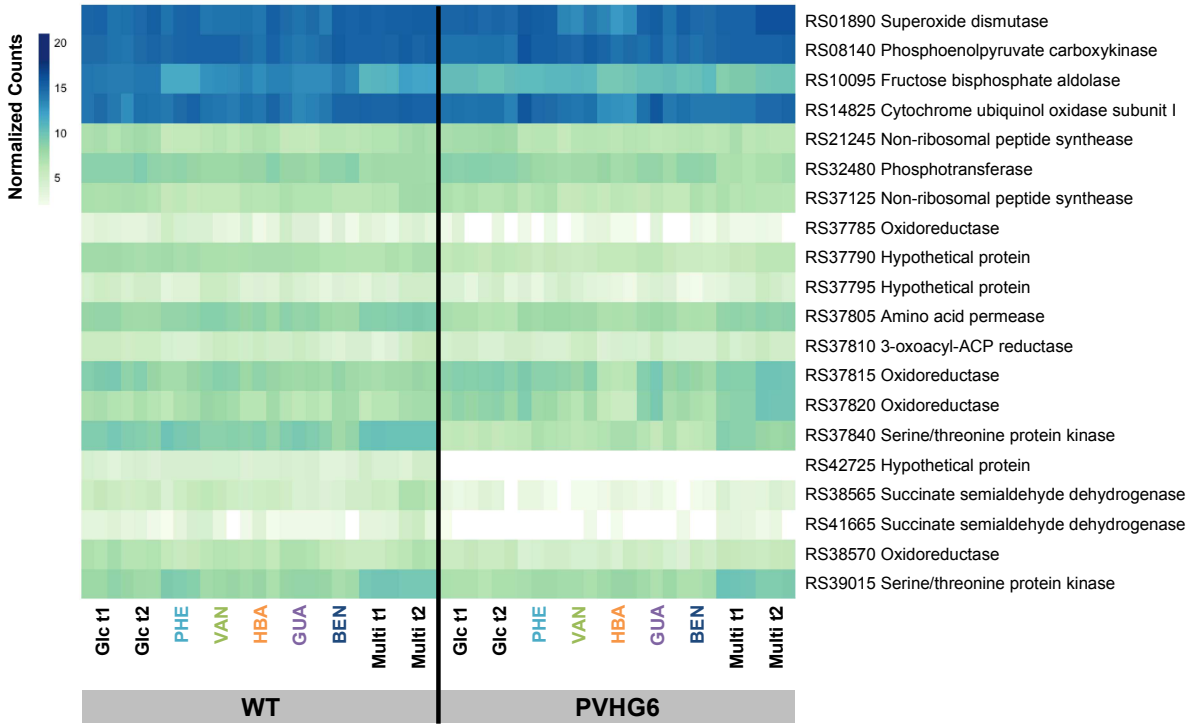


Figure 2-S21: Expression of genes with SNPs in PVHG6. Heat map of normalized gene expression counts for genes with SNPs in PVHG6. Glc = 1 g/L glucose, PHE = 0.5 g/L phenol, VAN = 0.5 g/L vanillate, HBA = 0.5 g/L 4-hydroxbenzoate, GUA = 0.5 g/L guaiacol, BEN = 0.5 g/L sodium benzoate, and Multi = 0.5 g/L of phenol, vanillate, 4-hydroxybenzoate, guaiacol, and sodium benzoate (2.5 g/L total aromatics). t_1 represents cells harvested at early exponential phase (10 hours for Glc; 20 hours for Multi), while t_2 represents cells harvested at mid-exponential phase (13 hours for Glc; 32 hours for Multi). Samples from individual compound growth conditions were harvested at mid-exponential phase (24 hours for WT PHE, 21 hours for PVHG6 PHE, 19 hours for WT and PVHG6 GUA, 11 hours for WT and PVHG6 HBA, 12 hours for WT and PVHG6 BEN, and 24 hours for WT and PVHG6 VAN). All transcriptomics experiments were performed in minimal medium A. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Gene codes are from the NCBI reference sequence NZ_CP003949.1.

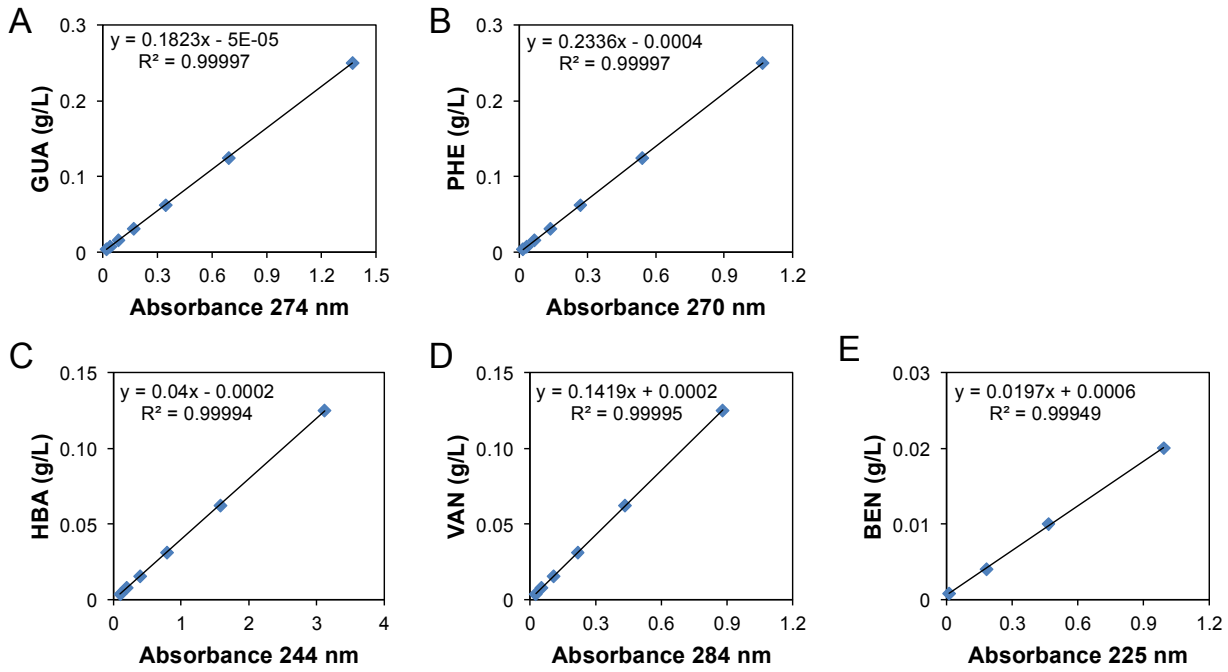


Figure 2-S22: Calibration curves for estimation of lignin model compound concentrations. Wavelength used for each calibration curve is listed on the x-axis of each plot, and the fitted correlation equation used to calculate concentrations are listed in each plot. A. Guaiacol calibration curve. B. Phenol calibration curve. C. 4-hydroxybenzoate calibration curve. D. Vanillate calibration curve. E. Sodium benzoate calibration curve. For A-D, measurements were performed by diluting 5 g/L stock solutions prepared in minimal medium A with no carbon or nitrogen sources and then measuring absorbance using a TECAN Infinite M200 Pro plate reader in a UV transparent 96 well plate. For E, dilutions were performed in the same way, but absorbance was measured using a Nanodrop 2000 with a quartz cuvette. See Materials and Methods for more information about measurement of lignin model compounds. PHE = phenol, GUA = guaiacol, VAN = vanillate, HBA = 4-hydroxybenzoate, BEN = sodium benzoate.

Table 2-S1: Adaptive evolution summary. Concentrations listed are final concentrations from adaptation. PCA = protocatechuate, GUA = guaiacol, PHE = phenol, HBA = 4-hydroxybenzoate, VAN = vanillate, BEN = sodium benzoate, and GLC = glucose. * No.: 3 = P, 14 = PV, 25 = PVH, 27 = PVHG, 32 = High glucose, and 33 = B.

No.*	# of subcultures	PCA (g/L)	GUA (g/L)	PHE (g/L)	HBA (g/L)	VAN (g/L)	BEN (g/L)	GLC (g/L)	Total aromatics (g/L)
1	40	4.58	-	-	-	-	-	-	4.58
2	83	-	1.67	-	-	-	-	-	1.67
3	80	-	-	1.50	-	-	-	-	1.50
4	60	-	-	-	3.09	-	-	-	3.09
5	48	-	-	-	-	2.06	-	-	2.06
6	56	2.85	0.47	-	-	-	-	-	3.32
7	63	3.32	-	0.55	-	-	-	-	3.87
8	60	2.63	-	-	0.88	-	-	-	3.51
9	62	3.24	-	-	-	1.08	-	-	4.32
10	62	-	0.64	0.64	-	-	-	-	1.27
11	42	-	0.30	-	1.06	-	-	-	1.36
12	59	-	1.10	-	-	1.10	-	-	2.19
13	58	-	-	1.05	1.05	-	-	-	2.09
14	71	-	-	1.16	-	1.16	-	-	2.32
15	56	-	-	-	1.41	1.41	-	-	2.82
16	64	1.34	0.22	0.22	-	-	-	-	1.78
17	64	2.91	0.48	-	0.48	-	-	-	3.88
18	74	3.18	0.53	-	-	0.53	-	-	4.25
19	60	2.14	-	0.36	0.36	-	-	-	2.86
20	74	3.19	-	0.53	-	0.53	-	-	4.25
21	74	2.26	-	-	0.75	0.38	-	-	3.39
22	101	-	0.59	0.59	1.19	-	-	-	2.37
23	70	-	0.34	0.34	-	0.69	-	-	1.37
24	79	-	0.66	-	1.33	1.33	-	-	3.32
25	67	-	-	0.49	0.98	0.98	-	-	2.44
26	54	1.36	0.23	0.23	0.45	-	-	-	2.26
27	104	-	0.49	0.49	0.98	0.98	-	-	2.95
28	50	1.54	-	0.26	0.51	0.51	-	-	2.82
29	42	2.99	0.50	0.50	-	1.00	-	-	4.98
30	57	1.52	0.25	-	0.51	0.51	-	-	2.78
31	60	1.51	0.25	0.25	0.50	0.50	-	-	3.02
32	72	-	-	-	-	-	-	419.0	0
33	80	-	-	-	-	-	12.58	-	12.58

Table 2-S2: Summary of comparative genomics of adapted strains. SNPs = single nucleotide polymorphisms; INDELs = insertions or deletions.

Sample	Total # of SNPs	# of Non-synonymous SNPs	# of Intergenic SNPs	# of Silent SNPs	# of Small INDELs	Plasmid 1 Deletions	Plasmid 2 Loss
P1	42	13	3	26	1	No	Yes
P2	45	16	3	26	1	No	Yes
P3	51	20	3	28	1	No	Yes
P4	46	15	3	28	1	No	Yes
P5	44	13	4	27	1	No	Yes
P6	50	16	4	30	1	No	Yes
PV1	29	9	1	19	0	No	Yes
PV2	29	9	1	19	1	No	Yes
PV3	36	9	3	24	0	No	Yes
PV4	28	8	1	19	0	No	Yes
PV5	30	10	1	19	0	No	Yes
PV6	34	11	2	21	1	No	Yes
PVH1	43	17	2	24	1	No	Yes
PVH2	36	14	2	20	1	No	Yes
PVH3	44	18	2	24	2	No	Yes
PVH5	45	19	2	24	2	No	Yes
PVH6	33	15	1	17	1	No	Yes
PVHG1	28	16	2	10	1	Yes	Yes
PVHG2	28	17	2	9	1	Yes	Yes
PVHG3	45	19	4	22	1	Yes	Yes
PVHG4	26	16	1	9	1	Yes	Yes
PVHG5	43	17	2	24	1	Yes	Yes
PVHG6	45	19	2	24	1	Yes	Yes
B1	3	2	0	1	0	No	Yes
B2	21	13	1	7	0	No	No
B3	10	6	2	2	0	No	No
B4	14	6	2	6	0	No	No
B5	46	20	3	23	0	Yes	Yes
B6	18	13	2	3	0	No	No
Glc1	30	17	2	11	0	Yes	No
Glc2	8	5	1	2	0	No	No
Glc3	28	14	3	11	0	Yes	No
Glc4	27	13	2	12	0	Yes	No
Glc5	7	4	1	2	0	No	No
Glc6	32	20	2	10	0	Yes	No

Table 2-S3: INDELs in adapted strains. INDELs were identified using Pilon variant calling, filtered with GATK, and verified with *de novo* assembly of the sequenced reads using SPAdes. Putative annotations are from NCBI reference sequence NC_CP003949.1. See Materials and Methods for more information. "All" indicates all sequenced strains isolated from the indicated adaptation condition. Strain names indicate adaptation condition from which the strain originated: P = phenol adaptation condition, PV = phenol and vanillate adaptation condition, PVH = phenol, vanillate, and 4-hydroxybenzoate adaptation condition, and PVHG = phenol, vanillate, 4-hydroxybenzoate, and guaiacol adaptation condition. Base Mean Expression is the average of the normalized counts from WT and PVHG6 transcriptomic data from all growth conditions.

Strains	Gene ID	Annotation	Type	Location	Base Mean Expression
All PVHG, All PVH	RS03610	penicillin binding protein	Insertion (1 bp)	Chromosome	8,100
PV2	RS18950	hypothetical protein	Insertion (5 bp)	Chromosome	334
All P	RS20545	hypothetical protein	Deletion (100 bp)	Chromosome	18
PV6	RS31910	helix-turn-helix transcriptional regulator	Insertion (1 bp)	Chromosome	753
PVH3, PVH5	RS40520	hypothetical protein	Insertion (1 bp)	Plasmid 8	116

Table 2-S4: Shared SNPs between strains isolated from the same adaptation condition. Genes with non-synonymous and intergenic SNPs that are shared between strains isolated from the same adaptation condition are shown. For intergenic SNPs, the nearest gene in each direction is listed. Some genes have more than one shared SNP.

Adaptation Condition	Gene ID	Annotation
P	RS01890/RS41735	superoxide dismutase/hypothetical protein
	RS14825	cytochrome ubiquinol oxidase subunit I
	RS21470	hypothetical protein
	RS37785/RS37790	SDR family mycofactocin-dependent oxidoreductase/hypothetical protein
	RS37790	hypothetical protein
	RS37795	hypothetical protein
	RS37805	amino acid permease
	RS37810	3-oxoacyl-ACP reductase
	RS37810/RS37815	3-oxoacyl-ACP reductase/LLM class flavin-dependent oxidoreductase
	RS37870	exoribonuclease II
PV	RS01890	superoxide dismutase
	RS10095	class II fructose-bisphosphate aldolase
	RS14825	cytochrome ubiquinol oxidase subunit I
	RS42075	hypothetical protein (pseudogene)
	RS37805	amino acid permease
	RS37810	3-oxoacyl-ACP reductase
	RS37810/RS37815	3-oxoacyl-ACP reductase/LLM class flavin-dependent oxidoreductase
PVH	RS00415	AMP-dependent synthetase
	RS02720	membrane protein
	RS02835	ammonia channel protein
	RS08140	phosphoenolpyruvate carboxykinase
	RS11540	4-hydroxy-2-oxovalerate aldolase
	RS14825	cytochrome ubiquinol oxidase subunit I
	RS25200	L-ectoine synthase
	RS27825	alpha/beta hydrolase
	RS37805	amino acid permease
	RS37810	3-oxoacyl-ACP reductase
	RS37810/RS37815	3-oxoacyl-ACP reductase/LLM class flavin-dependent oxidoreductase
PVHG	RS01890	superoxide dismutase
	RS11630	acyl-CoA dehydrogenase
	RS13070	non-ribosomal peptide synthetase
	RS14825	cytochrome ubiquinol oxidase subunit I
	RS22580	MFS transporter
	RS32480	phosphotransferase
	RS35615	peptidase
	RS37785/RS37790	SDR family mycofactocin-dependent oxidoreductase/hypotheical protein
	RS37790	hypothetical protein
	RS39015	serine/threonine-protein kinase PknK
B	RS03790	cold-shock protein
	RS08140	phosphoenolpyruvate carboxykinase
Glc	RS13475	glutamine--fructose-6-phosphate aminotransferase

RS15825	LysR family transcriptional regulator
RS24210	alpha-(1→6)-mannopyranosyltransferase A

Table 2-S5: Genes with non-synonymous SNPs or intergenic SNPs found across multiple adaptation experiments (including the glucose adaptation experiment). For intergenic SNPs, the nearest gene in each direction is listed. Some genes have more than one shared SNP. Putative annotations are from NCBI reference sequence NZ_CP003949.1. "All" indicates that all sequenced strains isolated from the indicated adaptation condition contain either non-synonymous SNPs or intergenic SNPs that could affect gene expression. Base Mean Expression is the average of the normalized counts from WT and PVHG6 transcriptomic data from all growth conditions.

Strains	Gene ID	Annotation	Location	Base Mean Expression
All P, All PV, PVH6, All PVHG	RS01890	superoxide dismutase	Chromosome	32,010
All B, All PVH	RS08140	phosphoenolpyruvate carboxykinase	Chromosome	30,198
P2, P3, P6, All PV	RS10095	class II fructose-bisphosphate aldolase	Chromosome	5,291
All P, All PV, All PVH, All PVHG	RS14825	cytochrome ubiquinol oxidase subunit I	Chromosome	23,629
PV5, B2, Glc6	RS21245	non-ribosomal peptide synthetase	Chromosome	110
PV3, All PVHG	RS32480	phosphotransferase	Chromosome	285
All P, PV3, B5, Glc1, Glc3, Glc4, Glc6	RS37785/ RS37790	intergenic (oxidoreductase/hypothetical protein)	Plasmid 1	8/103
All P, PVH3, All PVHG, B5, Glc1, Glc3, Glc4, Glc6	RS37790	hypothetical protein	Plasmid 1	103
All P, B5, Glc1, Glc3, Glc4, Glc6	RS37795	hypothetical protein	Plasmid 1	17
All P, All PV, All PVH, PVHG3, PVHG5, PVHG6, B4, B5, Glc1, Glc3, Glc4, Glc6	RS37805	amino acid permease	Plasmid 1	278
All P, All PV, All PVH, PVHG3, PVHG5, PVHG6, B5, Glc1, Glc3, Glc4, Glc6	RS37810	3-oxoacyl-ACP reductase	Plasmid 1	26
All P, All PV, All PVH, PVHG3, PVHG5, PVHG6, B5, Glc1, Glc3, Glc4, Glc6	RS37810/ RS37815	intergenic (3-oxoacyl-ACP reductase/oxidoreductase)	Plasmid 1	26/374
P2, P3, P4, P5,P6, PVHG2,PVHG4, B3, B4, Glc1, Glc3, Glc6	RS38565	NAD-dependent succinate-semialdehyde dehydrogenase (pseudogene)	Plasmid 1	24
P1, P2, P3, P4, P6, PV6, PVHG2, PVHG4, PVHG5, PVHG6, B3, B4, B5, B6, Glc1, Glc3, Glc6	RS41665	NAD-dependent succinate-semialdehyde dehydrogenase (pseudogene)	Plasmid 1	6
P5, PV3, PVH3, PVH5, B6	RS38570	FAD-binding oxidoreductase (pseudogene)	Plasmid 1	60
P2, P3, P4, P5, P6, All PVHG	RS39015	serine/threonine-protein kinase PknK	Plasmid 3	354

Table 2-S6: Phyre2 results for cytochrome ubiquinol oxidase I (RS14825). The native protein was modeled using the Protein Homology/analogY Recognition Engine (www.sbg.bio.ic.ac.uk/~phyre2/). 96% of residues were modeled at >90% confidence using the template protein 2YEV. Mutational sensitivity is defined as the likelihood of an amino acid change to affect function and is calculated on a scale of 0-9 (9 = the most sensitive).

Strain	SNP Change	AA Change	Predicted Mutational Sensitivity	Predicted Secondary Structure Location	Located in binding pocket?
All P, All PVHG	G to A	A272 to T	6	Transmembrane helix	Yes
All PV	T to C	I309 to T	5	Transmembrane helix	No
All PVH	T to G	F366 to V	8	Transmembrane helix	No

Table 2-S7: Phyre2 results for superoxide dismutase (RS01890). The native protein was modeled using the Protein Homology/analogY Recognition Engine (www.sbg.bio.ic.ac.uk/~phyre2/). 96% of residues were modeled at >90% confidence using the template protein 1GN4. Mutational sensitivity is defined as the likelihood of an amino acid change to affect function and is calculated on a scale of 0-9 (9 = the most sensitive).

Strain	SNP Change	AA Change	Predicted Mutational Sensitivity	Predicted Secondary Structure Location	Located in binding pocket?
All P	C to T	None, in regulatory region	Not applicable	Not applicable	Not applicable
All PV	A to C	V191 to G	8	Alpha helix	No
PVH6	A to G	V37 to G	7	Alpha helix	No
All PVHG	T to C	Y11 to C	8	None predicted	No

Table 2-S8: Growth characteristics of the WT and adapted strains using a mixture of lignin model compounds as carbon sources (2.5 g/L total aromatics). Growth experiments were performed in minimal medium A. Growth rates were calculated using at least four time points in the log phase. Statistical significance was determined with a threshold of $P < 0.05$ using a one mean, two-tailed Student's t-test. Growth rate and OD_{600} values are shown as the average of three biological replicates ± one standard deviation.

Strain	Growth rate (hr^{-1})	OD_{600} after 70 hr	Max OD_{600}	Growth rate significantly higher than WT?	OD_{600} after 70 hr significantly higher than WT?	Max OD_{600} significantly higher than WT?
WT	0.073 ± 0.001	0.54 ± 0.01	0.61 ± 0.01	not applicable	not applicable	not applicable
P1	0.084 ± 0.002	0.78 ± 0.02	0.81 ± 0.01	yes, $P = 5.3 \times 10^{-4}$	yes, $P = 2.1 \times 10^{-5}$	yes, $P = 2.2 \times 10^{-5}$
PV1	0.079 ± 0.002	0.82 ± 0.15	1.07 ± 0.13	yes, P = 0.0033	yes, P = 0.030	yes, P = 0.0034
PVH5	0.088 ± 0.001	0.55 ± 0.03	0.55 ± 0.03	yes, $P = 7.2 \times 10^{-5}$	no, P = 0.77	no, P = 0.053
PVHG6	0.088 ± 0.001	1.08 ± 0.02	1.08 ± 0.02	yes, $P = 2.6 \times 10^{-6}$	yes, $P = 3.8 \times 10^{-6}$	yes, $P = 5.1 \times 10^{-6}$
B2	0.085 ± 0.004	0.62 ± 0.03	0.69 ± 0.05	yes, P = 0.0086	yes, P = 0.019	no, P = 0.073

Table 2-S9: Growth characteristics of the WT and adapted strains using a mixture of lignin model compounds as carbon sources (3.0 g/L total aromatics). Growth experiments were performed in minimal medium A. Growth rates were calculated using at least four time points in the log phase. Statistical significance was determined with a threshold of $P < 0.05$ using a one mean, two-tailed Student's t-test. Growth rate and OD_{600} values are shown as the average of three biological replicates ± one standard deviation.

Strain	Growth rate (hr^{-1})	Max OD_{600}	Growth rate significantly higher than WT?	Max OD_{600} significantly higher than WT?
WT	0.059 ± 0.004	0.42 ± 0.02	not applicable	not applicable
P1	0.075 ± 0.001	0.73 ± 0.01	yes, P = 0.004	yes, $P = 3.8 \times 10^{-6}$
PV1	0.072 ± 0.004	0.74 ± 0.17	yes, P = 0.016	yes, P = 0.029
PVH5	0.073 ± 0.005	0.44 ± 0.04	yes, P = 0.021	no, P = 0.38
PVHG6	0.079 ± 0.001	0.77 ± 0.08	yes, P = 0.001	yes, P = 0.001
B2	0.077 ± 0.005	0.41 ± 0.04	yes, P = 0.011	no, P = 0.78

Table 2-S10: Half maximal inhibitory concentration (IC50) of individual lignin model compounds for WT and PVHG6 strains. Growth experiments were performed in minimal medium A. IC_{50} values were calculated using Prism (GraphPad Software, Inc.) with a four-parameter inhibitor response curve from OD_{600} measurements after 54 hours of growth. Statistical significance was determined with a threshold of $P < 0.05$ using a one mean, two-tailed Student's t-test. IC_{50} values are shown as the average of three biological replicates ± one standard deviation.

Compound	WT IC_{50} (g/L)	PVHG6 IC_{50} (g/L)	PVHG6 vs. WT difference (%)	PVHG6 significantly different from WT?
Phenol	1.07 ± 0.05	1.25 ± 0.01	17	yes, P = 0.003
Vanillate	2.12 ± 0.03	2.00 ± 0.05	-6	yes, P = 0.025
4-hydroxybenzoate	2.71 ± 0.02	2.73 ± 0.01	1	no, P = 0.157
Guaiacol	1.64 ± 0.04	1.90 ± 0.03	16	yes, P = 0.008
Sodium benzoate	7.41 ± 0.31	6.87 ± 0.16	-7	no, P = 0.055

Table 2-S11: Data from genome sequencing and RNA-Seq performed in this study. See the supplementary file at https://www.sciencedirect.com/science/article/pii/S1096717618300910 for Tables.

Tab 1: All SNPs in adapted strains organized by strain.
Tab 2: All SNPs in adapted strains organized by genome order.
Tab 3: All INDELs in adapted strains.
Tab 4: Summary of transcriptomic data for genes with mutations, β-ketoadipate pathway genes, funneling pathway genes, aromatic transporter genes, and C1 metabolism genes.
Tabs 5-26: Comparison of gene expression between samples using DESeq2.

Table 2-S12: Homology of RS34105 and RS34095 (upregulated in PVHG6 compared to WT) to putative 3-oxoadipyl-CoA thiolases. *R. opacus* PD630 proteins were compared to the *R. opacus* PD630 proteome and the *R. jostii* RHA1 proteome using NCBI blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The E value represents the number of hits expected by random chance when searching the database (the smaller value, the more significant match). The amino acid positive percent identity between each protein pair is reported as the number of identical or physically similar amino acids. If the protein has been characterized and named, the name is listed under notes.

***R. opacus* Gene code**	**Gene code**	**Amino Acid Positive % Identity**	**E value**	**Notes**
Pd630_RS34105	RHA1_ro02517	306/403 (75%)	0	CatF
	RHA1_ro01340	309/406 (76%)	0	PcaF
	RHA1_ro02863	220/417 (52%)	5e-81	PaaE
	Pd630_RS31575	303/401 (75%)	0	N/A
	Pd630_RS25375	309/406 (76%)	0	N/A
Pd630_RS34095	RHA1_ro02517	261/394 (66%)	3e-139	CatF
	RHA1_ro01340	259/403 (64%)	2e-127	PcaF
	RHA1_ro02863	229/412 (55%)	2e-70	PaaE
	Pd630_RS31575	262/396 (66%)	1e-135	N/A
	Pd630_RS25375	259/403 (64%)	1e-123	N/A

Table 2-S13: Homology of aromatic degradation enzymes in Rhodococcus opacus PD630 to enzymes in Rhodococcus jostii RHA1 and Rhodococcus opacus 1CP. Protein sequences from *R. opacus* PD630 were compared to the *R. jostii* RHA1 and *R. opacus* 1CP protein sequences using NCBI blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The E value represents the number of hits expected by random chance when searching the database (the smaller value, the more significant match). The amino acid positive percent identity between each protein pair is reported as the number of identical or physically similar amino acids. If the protein has been characterized and named, the name is listed under notes.

Fig. 2-5 genes (*R. jostii* RHA1)					
***R. opacus* PD630 gene code**	**Putative annotation**	***R. jostii* RHA1 gene code**	***R. jostii* % positive identity**	***R. jostii* E value**	**Notes**
RS31555	Two-component phenol hydroxylase, reductase component	RHA1_ro02513	189/190 (99%)	3E-139	PheA2
RS31560	Two-component phenol hydroxylase, oxygenase component	RHA1_ro02514	537/538 (99%)	0	PheA1
RS30765	Two-component phenol hydroxylase, reductase component	RHA1_ro02379	184/186 (98%)	5E-139	N/A
RS30770	Two-component phenol hydroxylase, oxygenase component	RHA1_ro02380	535/538 (99%)	0	N/A
RS02665	Putative vanillate demethylase, reductase component	RHA1_ro04163	308/319 (96%)	0	N/A
RS02670	Hypothetical protein	RHA1_ro04164	69/71 (97%)	3E-33	N/A
RS02675	Putative vanillate demethylase, oxygenase component	RHA1_ro04165	352/354 (99%)	0	N/A
RS31670	TetR family transcriptional regulator	RHA1_ro02538	208/208 (100%)	6E-150	N/A
RS31675	4-hydroxybenzoate 3-monooxygenase	RHA1_ro02539	390/392 (99%)	0	N/A
RS30780	Cytochrome P450	RHA1_ro02382	401/403 (99%)	0	N/A
RS30785	Putative cytochrome P450 reductase	RHA1_ro02383	320/329 (97%)	0	N/A

RS21470	Hypothetical protein	RHA1_ro02514	63/75 (84%)	3E-32	N/A
RS21475	Cytochrome P450	RHA1_ro08068	380/440 (86%)	0	N/A
RS21480	Putative cytochrome P450 reductase	RHA1_ro11319	284/323 (87%)	0	N/A
RS21485	Hypothetical protein	none	N/A	N/A	N/A
RS30790	Benzoate 1,2 dioxygenase subunit A	RHA1_ro02384	458/458 (100%)	0	BenA
RS30795	Benzoate 1,2 dioxygenase subunit B	RHA1_ro02385	167/169 (98%)	5E-124	BenB
RS30800	Benzoate 1,2 dioxygenase ferredoxin reductase	RHA1_ro02386	508/512 (99%)	0	BenC
RS30805	3,5-cyclohexadiene-1,2-diol-1-carboxylate dehydrogenase	RHA1_ro02387	262/264 (99%)	0	BenD

Fig. 2-5 genes (*R. opacus* 1CP)

***R. opacus* PD630 gene code**	**Putative annotation**	***R. opacus* 1CP gene code**	***R. opacus 1CP* % positive identity**	***R. opacus 1CP* E value**	**Notes**
RS31555	Two-component phenol hydroxylase, reductase component	R1CP_30420	188/190(98%)	7e-139	PheA2
RS31560	Two-component phenol hydroxylase, oxygenase component	R1CP_30425	536/538(99%)	0.0	PheA1
RS30765	Two-component phenol hydroxylase, reductase component	R1CP_29650	185/186(99%)	1e-139	N/A
RS30770	Two-component phenol hydroxylase, oxygenase component	R1CP_29655	534/538(99%)	0	N/A
RS31670	TetR family transcriptional regulator	R1CP_30540	207/208(99%)	3e-148	N/A
RS31675	4-hydroxybenzoate 3-monooxygenase	R1CP_30545	392/392(100%)	0	N/A
RS30790	Benzoate 1,2 dioxygenase subunit A	R1CP_29675	458/458 (100%)	0	BenA
RS30795	Benzoate 1,2 dioxygenase subunit B	R1CP_29680	169/169 (100%)	0	BenB
RS30800	Benzoate 1,2 dioxygenase ferredoxin reductase	R1CP_29685	508/512 (99%)	0	BenC
RS30805	3,5-cyclohexadiene-1,2-diol-1-carboxylate dehydrogenase	R1CP_29690	262/264 (99%)	0	BenD

Fig. 2-6 genes

***R. opacus* PD630 gene code**	**Putative annotation**	***R. jostii* RHA1 gene code**	***R. jostii* % Positive identity**	***R. jostii* E value**	**Notes**
RS25340	Succinyl-CoA:3-ketoacid-CoA transferase subunit A	RHA1_ro01333	212/213 (99%)	2E-155	PcaJ
RS25345	Succinyl-CoA:3-ketoacid-CoA transferase subunit B	RHA1_ro01334	249/249 (100%)	0	PcaI
RS25350	Protocatechuate 3,4-dioxygenase subunit A	RHA1_ro01335	230/233 (98%)	1E-173	PcaH
RS25355	Protocatechuate 3,4-dioxygenase subunit B	RHA1_ro01336	211/213 (99%)	2E-152	PcaG
RS25360	3-carboxy-muconate cycloisomerase	RHA1_ro01337	444/450 (98%)	0	PcaB
RS25365	3-oxoadipate enol-lactonase/4-carboxy-muconolactone decarboxylase	RHA1_ro01338	393/400 (98%)	0	PcaC
RS25370	IcIR family transcriptional regulator	RHA1_ro01339	264/265 (99%)	0	PcaR
RS25375	Putative 3-oxoadipyl-CoA thiolase	RHA1_ro01340	401/403 (99%)	0	PcaF
RS31565	Catechol 1,2-dioxygenase	RHA1_ro02515	278/282 (98%)	0	CatA2
RS31570	Putative 3-oxoadipate enol-lactonase	RHA1_ro02516	257/264 (97%)	0	MhpC
RS31575	Putative 3-oxoadipyl-CoA thiolase	RHA1_ro02517	402/410 (98%)	0	CatF
RS30720	ROK family transcriptional regulator	RHA1_ro02370	396/397 (99%)	0	N/A
RS30725	Muconolactone delta-isomerase	RHA1_ro02371	92/93 (98%)	3E-65	CatC
RS30730	Muconate cycloisomerase I	RHA1_ro02372	370/373 (99%)	0	CatB
RS30735	Catechol 1,2-dioxygenase	RHA1_ro02373	279/280 (99%)	0	CatA
RS30745	Intradiol ring-cleavage dioxygenase	RHA1_ro02374	253/256 (98%)	0	CatR

Table 2-S14: Lignin model compound concentrations for Fig. 2-2 seed cultures and colony purification of adapted cultures. For the strains not used for Fig. 2, the listed concentrations can be used for their seed cultures. PCA = protocatechuate, GUA = guaiacol, PHE = phenol, HBA = 4-hydroxybenzoate, VAN = vanillate, BEN = sodium benzoate, and GLC = glucose. * No.: 3 = P, 14 = PV, 25 = PVH, 27 = PVHG, 32 = High glucose, and 33 = B.

No.*	PCA (g/L)	GUA (g/L)	PHE (g/L)	HBA (g/L)	VAN (g/L)	BEN (g/L)	GLC (g/L)
1	2	-	-	-	-	-	-
2	-	0.5	-	-	-	-	-
3	-	-	0.5	-	-	-	-
4	-	-	-	1	-	-	-
5	-	-	-	-	1	-	-
6	1	0.3	-	-	-	-	-
7	1	-	0.3	-	-	-	-
8	1	-	-	0.5	-	-	-
9	1	-	-	-	0.5	-	-
10	-	0.3	0.3	-	-	-	-
11	-	0.3	-	0.5	-	-	-
12	-	0.3	-	-	0.5	-	-
13	-	-	0.3	0.5	-	-	-
14	-	-	0.3	-	0.5	-	-
15	-	-	-	0.5	0.5	-	-
16	0.75	0.2	0.2	-	-	-	-
17	0.75	0.2	-	0.3	-	-	-
18	0.75	0.2	-	-	0.3	-	-
19	0.75	-	0.2	0.3	-	-	-
20	0.75	-	0.2	-	0.3	-	-
21	0.75	-	-	0.3	0.3	-	-
22	-	0.2	0.2	0.3	-	-	-
23	-	0.2	0.2	-	0.3	-	-
24	-	0.2	-	0.3	0.3	-	-
25	-	-	0.2	0.3	0.3	-	-
26	0.5	0.1	0.1	0.2	-	-	-
27	-	0.1	0.1	0.2	0.2	-	-
28	0.5	-	0.1	0.2	0.2	-	-
29	0.5	0.1	0.1	-	0.2	-	-
30	0.5	0.1	-	0.2	0.2	-	-
31	0.4	0.1	0.1	0.1	0.1	-	-
32	-	-	-	-	-	-	100
33	-	-	-	-	-	3	-

Table 2-S15: Lignin model compound concentrations for Fig. 2-2 subcultures. For the strains not used for Fig. 2, the listed concentrations can be used for their subcultures. PCA = protocatechuate, GUA = guaiacol, PHE = phenol, HBA = 4-hydroxybenzoate, VAN = vanillate, BEN = sodium benzoate, and GLC = glucose. * No.: 3 = P, 14 = PV, 25 = PVH, 27 = PVHG, 32 = High glucose, and 33 = B.

No.*	PCA (g/L)	GUA (g/L)	PHE (g/L)	HBA (g/L)	VAN (g/L)	BEN (g/L)	GLC (g/L)
1	2.29	-	-	-	-	-	-
2	-	0.84	-	-	-	-	-
3	-	-	0.75	-	-	-	-
4	-	-	-	1.54	-	-	-
5	-	-	-	-	1.03	-	-
6	1.42	0.24	-	-	-	-	-
7	1.66	-	0.28	-	-	-	-
8	1.32	-	-	0.44	-	-	-
9	1.62	-	-	-	0.54	-	-
10	-	0.32	0.32	-	-	-	-
11	-	0.15	-	0.53	-	-	-
12	-	0.55	-	-	0.55	-	-
13	-	-	0.52	0.52	-	-	-
14	-	-	0.58	-	0.58	-	-
15	-	-	-	0.70	0.70	-	-
16	0.67	0.11	0.11	-	-	-	-
17	1.45	0.24	-	0.24	-	-	-
18	1.59	0.27	-	-	0.27	-	-
19	1.07	-	0.18	0.18	-	-	-
20	1.59	-	0.27	-	0.27	-	-
21	1.13	-	-	0.38	0.19	-	-
22	-	0.30	0.30	0.59	-	-	-
23	-	0.17	0.17	-	0.34	-	-
24	-	0.33	-	0.66	0.66	-	-
25	-	-	0.24	0.49	0.49	-	-
26	0.68	0.11	0.11	0.23	-	-	-
27	-	0.25	0.25	0.49	0.49	-	-
28	0.77	-	0.13	0.26	0.26	-	-
29	1.49	0.25	0.25	-	0.50	-	-
30	0.76	0.13	-	0.25	0.25	-	-
31	0.75	0.13	0.13	0.25	0.25	-	-
32	-	-	-	-	-	-	210
33	-	-	-	-	-	6.29	-

Table 2-S16: Summary of heterologous plasmids used in this work.

Plasmid Name	Function	Backbone	Antibiotic	Length (bp)	Source
pDD120	pConstitutive.Che9c60.Che9c61	pB264	Kanamycin	6247	Delorenzo et al. 2018 [139]
pDD163	RS30730 knockout vector	Integrative/p15a	Hygromycin B	3268	This study
pDD174	RS25360 knockout vector	Integrative/p15a	Gentamicin	4249	This study
pDD175	RS30810 knockout vector	Integrative/p15a	Hygromycin B	3373	This study
pDD176	RS33590 knockout vector	Integrative/p15a	Hygromycin B	3336	This study
pDD177	RS031355 knockout vector	Integrative/p15a	Hygromycin B	3327	This study

Table 2-S17: Summary of strains constructed using heterologous plasmids in Table 2-S16.

Strain name	Genus	Species	Strain	Plasmid contained or used	Corresponding figure
DMD155	*Escherichia*	*coli*	DH10B	pDD120	N/A
DMD250	*Escherichia*	*coli*	DH10B	pDD163	N/A
DMD251	*Rhodococcus*	*opacus*	PD630	pDD163	Fig. 6C
DMD258	*Escherichia*	*coli*	DH10B	pDD177	N/A
DMD270	*Escherichia*	*coli*	DH10B	pDD174	N/A
DMD274	*Escherichia*	*coli*	DH10B	pDD175	N/A
DMD275	*Escherichia*	*coli*	DH10B	pDD176	N/A
DMD281	*Rhodococcus*	*opacus*	PD630	pDD174	Fig. 6D
DMD282	*Rhodococcus*	*opacus*	PD630	pDD177	Fig. S18
DMD283	*Rhodococcus*	*opacus*	PD630	pDD175	Fig. S18
DMD284	*Rhodococcus*	*opacus*	PD630	pDD176	Fig. S18

2.7 Acknowledgments

The authors thank members of the Dantas, Moon, and Foston labs for scientific discussions and critical review of the manuscript, and Jessica Hoisington-Lopez, Eric Martin, and Brian Koebbe in the Edison Family Center for Genome Sciences and Systems Biology at the Washington University in St. Louis School of Medicine for Illumina sequencing and high-throughput computing support. This work was supported by the U.S. Department of Energy [Grant nos. DE-SC0012705 and DE-SC0018324 to G.D., T.S.M., and M.F.].

Chapter 3

The microbiome and resistome of chimpanzees, gorillas, and humans across lifestyle and geography

Collaboration Statement

This work was performed in a collaboration between Tayte P. Campbell, Xiaoqing Sun, Vishal H. Patel, Crickette Sanz, David Morgan, and Gautam Dantas. T.P.C., X.S., G.D., D.M., and C.S. designed the study and experiments. D.M. and C.S. collected samples. T.P.C., X.S., and V.H.P. performed the experiments. T.P.C. analyzed the results. T.P.C, G.D., D.M., and C.S. wrote the paper.

3.1 Abstract

The gut microbiome is generally stable but can vary across differences in host lifestyle, geography, and host species. By comparing closely related host species across varying locations and geography we can evaluate the relative contributions of these factors in structuring the composition and functions of the microbiome. Here we show that the gut microbiome and resistome of great apes and humans are more related by host lifestyle than geography. Using

shotgun and functional metagenomics, we show that captive chimpanzees and gorillas have an enrichment of microbial genera commonly found in non-Westernized humans despite living in close geographic proximity to Westernized humans. Captive apes also maintain higher richness and abundance of antibiotic resistance genes compared to wild apes. Finally, by comparing our study cohorts to human gut microbiomes from a diverse range of environments and lifestyles, and to other wild and captive gorilla cohorts, we find that the influence of host lifestyle is robust to various geographic locations.

3.2 Introduction

From an anthropocentric view, the microbiomes of great apes provide a view to our evolutionary past [140-142]. Gut microbial communities are generally stable, but they can be altered in accordance with several factors including host lifestyle [140, 143-151], host species [141, 142, 152], and geography [153-155]. By comparing microbial variation in wild and captive great apes, we can approximate changes that occurred in human modernization as human microbiomes were influenced by changes in lifestyle, including an increasingly Westernized diet and increased antibiotic usage [140]. Such comparisons are also important for understanding the impact of humans and their activities on great ape health. Anthropogenic threats such as poaching, habitat encroachment, and diseases are the primary causes of population decrease in these endangered Hominidae species, and the risk to great ape populations of exposure to human activities continues to increase [154, 156-159].

The taxonomic composition of microbiomes of captive primates from the United States, Southeast Asia, and Costa Rica have been shown to cluster with those of geographically separated non-Westernized humans [143]. By synchronous sampling of a non-Westernized human cohort that is sympatric with wild apes, we can test if their gut microbiomes vary predominantly by host

species, lifestyle, or geographic distance [140, 160]. These findings can help guide conservation efforts in endangered animals aimed at promoting microbiome configurations that are linked with optimal host health and metabolism [161]. Such insight will not only inform the well-being of captive populations but can lead to improved conservation outcomes in rehabilitation and reintroduction programs [162].

Antibiotic exposure represents a key lifestyle-associated factor which can acutely and persistently alter the gut microbiome [145, 151, 163]. Because of their ability to kill or inhibit the growth of virtually all microbes, antibiotic treatments provide strong selection for enrichment of antibiotic resistance genes (ARGs), selecting for and leading to an increasingly antibiotic resistant bacterial composition [164, 165]. However, ARGs can also be found in bacteria from pristine environments naïve to human-produced antibiotics, most likely due to the selection pressure provided by natural antibiotic secretion from environmental bacteria and fungi over long evolutionary time-scales [145, 166]. The discovery, and often indiscriminate usage, of antibiotics in human and agricultural settings over the past century has selected for dramatic enrichment and transmission of ARGs between environmental, commensal, and pathogenic microbes from diverse habitats [145, 167]. Proactive identification of ARGs in environmental and host-associated microbiomes allows for their characterization, tracking, and potential inhibition before they can be exchanged with pathogens and found in the clinic [163, 168, 169]. Characterization of the antibiotic resistome of wild and captive non-human primates can help approximate how the human resistome may have changed in the antibiotic era, as well as assess the impact of human-contact and captivity on the primate gut resistome [145, 151].

Here we compare the fecal microbiome taxonomic composition, microbial functional pathways, and antibiotic resistome of wild and captive chimpanzees and gorillas, as well as

humans living in the same geographical region as the wild apes. We also compare these cohorts with human cohorts from around the world representing diverse lifestyles and geographies [144, 146, 147, 163, 170]. We hypothesize that the gut microbiomes, functional pathways, and resistomes of sympatric humans and apes will be similar to each other since sharing of gut microbiota between humans and wild and domestic animals living in geographic proximity has been shown previously [154, 155, 171]. We further hypothesize that captive chimpanzees and gorillas from zoos in the United States will be most similar to Westernized human microbiomes because of human exposure and antibiotic treatment.

In this study we used high-throughput metagenomic sequencing to characterize the gut microbiota and resistomes of 18 wild central chimpanzees (*Pan troglodytes troglodytes*) and 28 wild western lowland gorillas (*Gorilla gorilla gorilla*) from Nouabalé-Ndoki National Park in the Republic of the Congo. These wild animals live in a remote area of the park that was completely naïve to human impact before 2003 and identified as the ape population that was least disturbed by humans in 2014 [172, 173]. We also collected samples from 81 humans living just outside of the park, including some individuals that work within the park, in the Republic of the Congo. We compared these cohorts to 18 chimpanzees and 15 western lowland gorillas from the St. Louis Zoo (St. Louis, MO) and the Lincoln Park Zoo (Chicago, IL) in the United States. To tease apart the contribution of lifestyle, location, and host species, our study includes two lifestyles (wild hunter-gatherer and non-Western human), two locations (USA and the Republic of the Congo), and three host species (human, chimpanzee, and gorilla). To determine if our results are robust to further differences in lifestyles and locations, we compared the fecal microbiomes from our samples to published hunter-gatherer, rural agriculturalist, and urban human cohorts from the USA [146, 174], Peru [146, 163], El Salvador [163], Malawi [144], Tanzania [147, 170], and Venezuela [144], and

to published wild and captive gorilla cohorts from the Republic of the Congo, the USA, France, and Switzerland [140, 175].

3.3 Results & Discussion

3.3.1 Taxonomic composition of the gut microbiome was predominantly influenced by lifestyle

To test our hypothesis that the microbiomes of sympatric hosts will be the most similar, we sequenced microbial metagenomic DNA from the fecal samples of our wild chimpanzee and gorilla cohorts from the Republic of the Congo, humans living in close proximity, and captive chimpanzees and gorillas from zoos in the USA. Overall, we generated 10.4 million sequences of the V4 region of the 16S ribosomal RNA gene from 160 samples, which we analyzed as exact amplicon sequence variants (ASVs). Through Principal Coordinates Analysis (PCoA), we found that captive chimpanzees and gorillas clustered more closely with the non-Westernized humans than their wild counterparts, indicating more human-like microbiomes (Figures 3-1A, 3-S1A). Despite living in close geographic proximity, the humans, wild chimpanzees, and wild gorillas live vastly different lifestyles as characterized by diet and antibiotic usage. Wild chimpanzees and gorillas both eat natural plants and have some overlap in their diets, but chimpanzees are frugivores while gorillas are predominantly folivores [176-178]. The humans in this study are rural agriculturists and eat non-Westernized diets, high in plant fiber and low in animal protein, and are exposed to antibiotic treatments that can further alter the gut microbiome [149, 179, 180]. The captive apes have similar diets and a similar degree of antibiotic treatment as non-Westernized humans which most likely drive the similarities in humans and captive apes [181]. This provides evidence that host lifestyle is more influential to primate gut microbiome composition than either

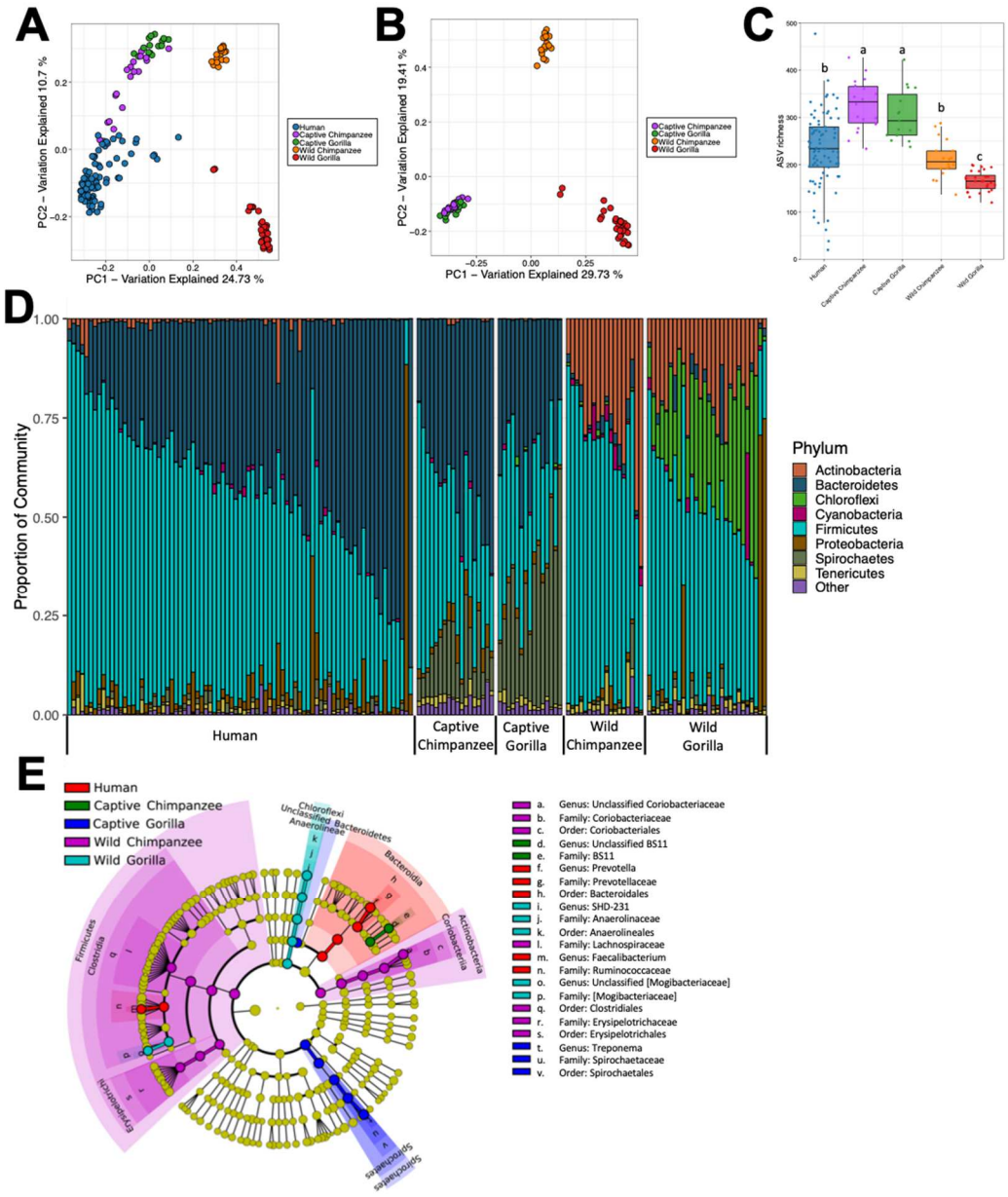


Figure 3-1: Taxonomic composition of the human, chimpanzee, and gorilla gut microbiomes. (A) Principal-coordinate analysis (PCoA) plot of Bray-Curtis 16S metagenomic profiles from individual human, chimpanzee, and gorilla fecal samples. (B) PCoA plot of Bray-Curtis 16S metagenomic profiles from individual chimpanzee and gorilla fecal samples. (C) Boxplot of ASV richness in host microbiomes (Kruskal-Wallis sum-rank test $P < 3.8e\text{-}14$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 7.7e\text{-}4$). (D) Relative abundances of major bacteria phyla for all individuals. (E) Cladogram of discriminatory taxa identified in human, chimpanzee, and gorilla microbiome 16S data sets. Clades significantly enriched in human (red nodes), captive chimpanzee (green nodes), captive gorilla (blue nodes), wild chimpanzee (purple nodes), and wild gorilla (light blue nodes) metagenomes are indicated (LEfSe; linear discriminant analysis (LDA) log score >4.5, $P = 0.05$).

geography or host species. These findings are consistent with previous comparisons between captive and wild douc and howler monkeys as well as comparisons between non-Westernized and Westernized humans [143, 180].

The captive ape microbiomes all clustered more closely together than wild apes despite being housed in two zoos ~500 km apart (Figure 3-1B). This finding is in contrast to a previous study where the gut microbiomes of wild sympatric chimpanzees and gorillas were more similar than wild allopatric apes [153]. Comparisons of sympatric and allopatric wild apes may be confounded by the underlying changes in plants that are available for consumption in the different locations. In contrast, the relatively limited diversity in human agriculture limits the diets for captive apes and non-Westernized humans influencing their microbiomes to become more similar [182]. Host species appeared to have some effect on the gut microbiome since wild and captive chimpanzee microbiomes clustered more closely to humans than the respective wild and captive gorillas. Chimpanzees are evolutionarily more closely related to humans than gorillas are, and that relatedness is maintained in their gut microbiome in these findings [141, 142].

To identify differences in alpha diversity, we compared the species richness and Shannon diversity of microbiomes in our cohort. We found that captive apes in the USA had higher richness of ASVs (Figure 3-1C) following a geographic pattern, while humans and chimpanzees had higher Shannon diversity than gorillas following a pattern of being influenced by host species (Figures 3-1C, 3-S1B). Shannon diversity takes into account species evenness, with lower Shannon values indicating either decreased richness, decreased evenness, or both. Since captive apes have higher species richness but similar Shannon values to wild apes, their microbiomes must contain a high number of rare ASVs. Previous studies have reported that captivity is typically associated with a decrease in microbiome richness, but in light of our findings there appears to be more nuance to

that pattern depending on the population studied [143, 183]. In support of this, a previous study found no significant difference in the Shannon diversity of wild and captive gorillas [175].

We used linear discriminant analysis effect size (LEfSe) [184] analysis to identify specific taxa that are enriched in each host microbiome and found that captivity is associated with a higher abundance of *Prevotella* (Bacteroidetes) and *Treponema* (Spirochaetes) for both chimpanzees and gorillas (Figures 3-1D,E, 3-S1D,E). *Prevotella* was also found in high abundance in the Congolese human microbiomes we sampled. Increased colonization of *Prevotella* and *Treponema* in captive animal microbiomes and non-Westernized humans has been observed previously and is often attributed to host diet [143, 149]. Species within the genus *Prevotella* are capable of plant polysaccharide degradation, exhibiting ß-D-xylanase and ß-xylosidase activities which are associated with hemicellulose degradation [143, 149]. Increased abundance of *Prevotella* in captive apes suggests that their increased abundance may be caused by differences in the composition of dietary polysaccharides compared to wild apes [143]. We were able to identify three *Treponema* ASVs at the species level: *Treponema succinifaciens, Treponema bryantii*, and *Treponema berlinense*. These species of *Treponema* are considered non-pathogenic commensals that metabolize complex polysaccharides found in unprocessed plants [185, 186]. Commensal *Treponema* species have been found in high abundance in human populations with non-Westernized lifestyles, including Hadza hunter-gatherers in Tanzania [148], Matses hunter-gatherers in the Peruvian Amazon [146], humans practicing traditional agriculture in the Tunapuco community of the Peruvian Andean highlands [146], and BaAka pygmies in the Central African Republic [149]. Due to its presence in non-Westernized human populations and absence in Westernized humans, *Treponema* has been postulated to be a member of the ancestral human gut microbiome [149]. Low levels of *Treponema* present in our wild ape samples support this

hypothesis since they indicate that this genus is naturally shared in our evolutionarily nearest living relatives. The much higher abundance in captive apes compared to wild apes indicates that *Treponema* species may be optimized to metabolize plant polysaccharides from the human diet [146].

In contrast, wild chimpanzees were distinguished by higher relative recovery of *Coriobacteriaceae* (Actinobacteria) and two families from the phylum Firmicutes: *Erysipelotrichaceae* and *Lachnospiraceae* (Figures 3-1E, 3-S1D). Wild gorillas had higher relative recovery of *Mogibacteriacea* (Firmicutes) and *SHD-231* (Chloroflexi), an under-characterized bacterial genus found previously in wild western lowland gorilla microbiomes [140] (Figures 3-1D, 3-S1E). *SHD-231* appears to be found only within the gorilla microbiome, indicating that it may have co-evolved with the gorilla microbiome.

3.3.2 Functional pathways in the microbiome are relatively robust to changes in individual taxa

To determine how differences in microbiome taxonomy altered functional potential, we performed whole metagenome shotgun sequencing on all 160 fecal samples, rarefied the resulting sequences to four million paired-end reads per sample, and annotated functional pathways using HUMAnN2 [35]. We found that lifestyle had less influence on functional pathways in the microbiome than on taxa (Figures 3-2A, 3-S2A,B). Functional pathways may be robust to changes in taxa since many pathways are redundant across microbial taxa[35]. Using LEfSe, we identified several discriminatory pathways that showed varying patterns of association with lifestyle and host species (Figures 3-2B, 3-S2C). Captive and wild apes had higher relative recovery of branched amino acid synthesis pathways (including isoleucine and valine) compared to humans, with decreasing relative abundance from wild to captive apes and finally dropping further in humans

(Figure 3-2B). This indicates that the non-Westernized lifestyle is associated with decreasing branched amino acid synthesis but the large difference in ape and human relative abundance indicates that host species also influences these changes. Congolese humans had the highest levels of lysine and aromatic amino acid synthesis with relative abundance decreasing along both the gradient of non-Westernized human to wild ape lifestyles, and host species relatedness with relative abundance decreasing as the relatedness to humans decreased.

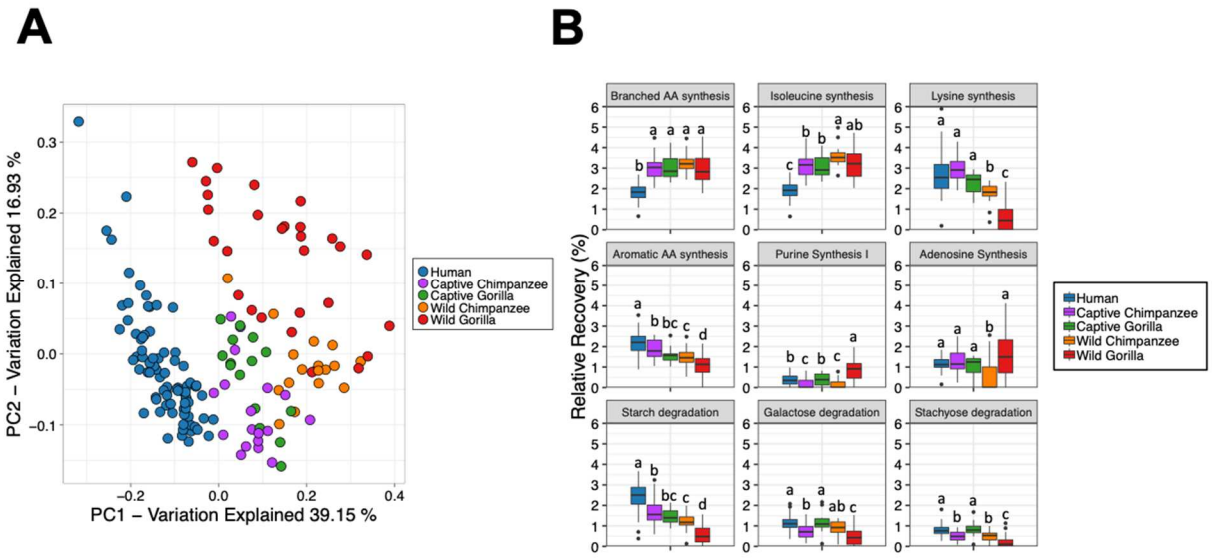


Figure 3-2: Functional pathway analysis of gut microbiota. (A) PCoA plot of Bray-Curtis distances of the relative abundance of functional pathways from HUMAnN2 results. (B) Relative abundance of functional pathways involved in amino acid synthesis, peptidoglycan synthesis, and carbohydrate degradation (Kruskal-Wallis sum-rank test for all plots $P \leq 3.2E\text{-}9$; Wilcoxon rank sum test with Benjamini-Hochberg correction, for significant differences in all plots $P \leq 0.047$).

The superpathway of purine synthesis I and adenosine synthesis was highest in the wild gorillas and may be associated with the *SHD-231* family, as a previous study also found increased nucleotide sugar biosynthesis and purine nucleotide salvage pathways in gorilla microbiomes dominated by *SHD-231* [140]. We next investigated carbohydrate degradation pathways and found that Congolese human microbiomes were enriched for the degradation of starch, galactose, and stachyose. Again, we saw decreasing relative abundance of these pathways as host lifestyle

becomes more different from the non-Westernized human lifestyle. For galactose and stachyose degradation, host species influenced relative recovery as captive and wild chimpanzees had no difference in relative recovery, but wild gorillas had decreased abundances compared to captive gorillas. The relative abundance of the starch degradation pathway shows a clear pattern of being influenced by host lifestyle. Starch consumption is the main source of calories in humans, while wild chimpanzees and gorillas consume low starch diets [181, 187]. Starch and carbohydrate degradation have been linked with the presence of *Treponema* and *Prevotella*. These genera may be the predominant taxa providing these functions in our cohorts since they are also enriched in humans and captive apes in this study [140, 143].

3.3.3 Microbial resistome is primarily influenced by host lifestyle

To probe for antibiotic resistance genes (ARGs) that were putatively unique to our cohorts, we used functional metagenomics, a method where metagenomic DNA fragments are cloned into expression vectors, transformed into *E. coli*, and the resultant transformants are tested for phenotypic resistance using antibiotic selection [12, 151, 163]. From our 160 samples, we created 16 functional libraries containing 50-123 gigabases of total insert DNA and screened them on 15 different antibiotics or antibiotic combinations resulting in the annotation of 332 ARGs (Figure 3-3A, Table 3-S1). We combined these resulting protein sequences with ARG protein sequences from the Comprehensive Antibiotic Resistance Database (CARD) [188], and used ShortBRED [36] to create unique markers for ARG identification in the 640 million shotgun metagenomic reads (4 million per sample) originating from ape and human fecal DNA. We found that captive apes clustered more closely to humans than wild apes for both abundance and presence of ARGs (Figures 3-3A,B, 3-S3D,E) and, similar to our taxonomic results, captive ape resistomes clustered

more closely together than wild apes (Figures 3-3B,C). This provides evidence that host lifestyle influences the gut resistome more than geography or host species for the studied primates.

Aminoglycoside resistance genes represented more than 50% of the wild ape gut resistomes (Figure 3-3C). Aminoglycosides are produced naturally in the environment by several soil bacteria and inhibit protein synthesis by binding to the ribosome [189]. The specific aminoglycoside resistance genes we found enriched in wild apes were identified as AAC(3)-VIIa, an acetyltransferase that inactivates aminoglycosides [189], and *FmrO* ribosomal RNA methyltransferases [190] that were identified from our functional selections. Ribosomal methyltransferases provide aminoglycoside resistance by methylation of the ribosome that prevents aminoglycoside binding and thus resistance [191]. The overall abundance and richness of ARGs in wild apes was extremely low compared to humans and captive apes (Figure 3-3D). In captive apes and humans, tetracycline resistance genes were the most abundant with beta-lactams also making up a large proportion of the resistome (Figure 3-3C). Tetracyclines and beta-lactams are both broad-spectrum antibiotics and are popular first-line antibiotics in human infectious disease treatment [192]. Tetracycline resistance in the human and captive ape cohorts was conferred predominantly by six different ribosomal protection genes identified through tetracycline functional selections and beta-lactam resistance was dominated by high recovery of CfxA4, a beta-lactamase. Captive apes had up to ~34-fold higher abundance and up to ~5-fold higher richness of all ARGs compared to wild apes (Figures 3-3D,E). The most parsimonious explanation for the higher observed burden of ARGs in captive apes compared to wild apes is the higher selection pressure from antibiotic treatment of the captive apes. We had limited access to antibiotic treatment records for captive apes but the records we do have indicate that captive apes received on average 1.7 antibiotic treatments per year providing ample selection for ARGs.

Congolese human microbiomes had the highest abundance and richness of ARGs with ~3-fold higher abundance and ~2-fold higher richness compared to the next highest cohort (Figures 3-3D,E). Host lifestyle appears to be the driving force for ARG abundance and richness in these analyses.

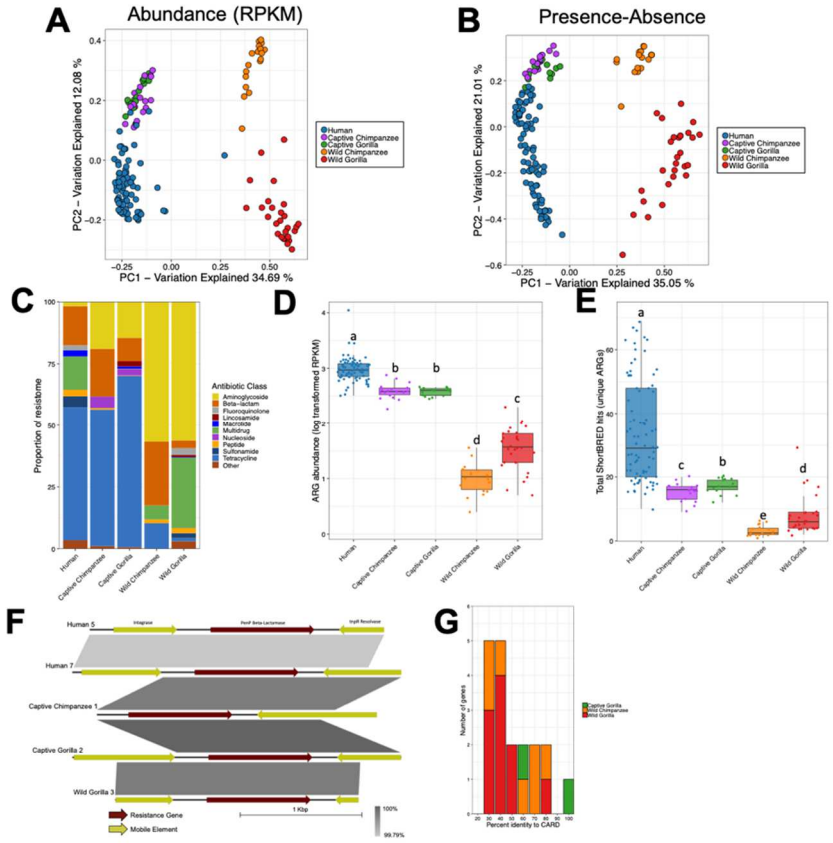


Figure 3-3: The antibiotic resistome of humans, chimpanzees, and gorillas. (A) PCoA plot using Bray-Curtis distances of ARG abundance normalized with RPKM from ShortBRED results of human, chimpanzee, and gorilla resistomes. (B) PCoA plot using Bray-Curtis distances of ARG richness (presence/absence) from ShortBRED results of human, chimpanzee, and gorilla resistomes. (C) Relative abundances of major classes of antibiotic resistance averaged by host cohort. (D) Sum of all ARG marker abundances per metagenome based on markers generated from the CARD database and functional selections, expressed in log transformed RPKM (Kruskal-Wallis sum-rank test $P < 2.2e\text{-}16$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 1.2e\text{-}5$). (E) Resistome richness per metagenome (Kruskal-Wallis sum-rank test $P < 2.2e\text{-}16$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 0.049$). (F) Alignment of an ARG containing DNA fragment with high similarity across host cohorts obtained from functional metagenomic selections. (G) Histogram of amino acid identity to known ARGs for aminoglycoside resistance genes identified through functional metagenomic selections. Colors indicate from which host libraries each gene was isolated.

3.3.4 Functional metagenomic selections identified novel antibiotic resistance genes

One major advantage of functional metagenomic analysis is the ability to identify novel ARGs from metagenomic DNA without the need for high sequence identity to known ARGs [12, 151, 163]. This allows for the characterization and surveillance of identified ARGs before they become clinically disseminated [163, 167, 193]. Since wild ape resistomes were dominated by aminoglycoside resistance, we explored proteins annotated as aminoglycoside resistance from our functional selections on ape microbiomes. From these selections, we identified 19 aminoglycoside resistance methyltransferases that belonged to the *FmrO* family. When compared to known ARGs in the CARD database, we found that they all aligned to 16S rRNA methyltransferases, such as *rmtC*, *rmtD*, *rmtF*, *rmtH*, or *sgm*. The percent similarity of these genes to known ARGs ranged from 29-98%, indicating that some of these genes have not been identified previously (Figure 3-3G).

Additionally, we found a total of nine contigs from functional metagenomic selections containing ARGs that were shared either across host species or captivity status demonstrating that Congolese human microbiomes harbor some of the same ARGs as wild and captive ape microbiomes (Figures 3-3F, 3-S4). One ~2 kb contig that was particularly widespread between different hosts contained a *PenP* beta-lactamase located between two mobile elements, an integrase and a *tnpR* resolvase, indicating a high potential for horizontal gene transfer [194, 195]. Highly similar copies (>99% similarity) of this *PenP*-containing contig were found in functional libraries created from every research cohort in this study except for wild chimpanzees. This gene may be widespread either because it is contained in a widespread bacterial taxon, has spread to several bacterial taxa through horizontal gene transfer, or a combination of these factors. Out of

the nine *PenP*-containing contigs identified, three contained ARGs co-localized with mobile elements. Identification of ARGs in captive and wild apes indicates that ape microbiomes may serve as a zoonotic reservoir of antibiotic resistance that may be exchanged with human associated microbes [169]. Importantly, identification of similar ARGs in human and ape populations also indicates that the human microbiome may serve as a reservoir for ARG acquisition by great ape microbiomes. Transmission of microorganisms between humans and great apes in both directions is well documented, with several pathogens exhibiting zoonotic or anthropozoonotic patterns [154, 156-159, 171, 183, 196, 197]. Transmission of microorganisms also results in transfer of the ARGs that those organisms contain. Identification and characterization of reservoirs of ARGs in humans and other closely related Hominidae species, including wild and captive populations, allows for a proactive approach to combat antibiotic resistance via surveillance and targeted infection control measures [198-200]. These transmission events are likely to rise as humans and great apes are increasingly brought into contact with each other through ecotourism, human population growth near conservation areas, and civil unrest [154, 156, 183, 201, 202].

We also identified an ARG conferring resistance to colistin, an antibiotic of last-resort, from a Congolese human derived functional metagenomic library (Figure 3-S3A). The identified contig contained a gene annotated as a member of the type 2 phosphatidic acid phosphatase (PAP2) superfamily that, when isolated and expressed in a pZE21 plasmid backbone, conferred a minimal inhibitory concentration of 4 μg/ml to colistin in *E. coli*. This represents an ~85-fold increase in resistance compared to the empty vector control (Figure 3-S5A). An MIC of 4 μg/ml exceeds the 2 μg/ml resistant/susceptible breakpoint in *Enterobacteriaceae* according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST), indicating this level of resistance could have implications if this gene were transmitted to human pathogens. Colistin works by

targeting the lipid A portion of lipopolysaccharide leading to disruption of the outer cell membrane and ultimately cell death [203]. Resistance is typically conferred by modification of lipid A by addition of chemical moieties to LPS, preventing colistin binding. For example, a clinically relevant colistin resistance gene, *mcr*-1, provides colistin resistance by adding a phosphoethanolamine moiety to LPS [198]. We predict that the PAP2 gene from this study either alters LPS by adding a chemical moiety onto LPS or by altering the composition of the cell membrane to include less LPS.

Using BLAST, we identified two proteins similar to the PAP2 gene product in the NCBI non-redundant protein database. One was discovered from published functional metagenomic selections on colistin [163] (98.62% identity and 76% coverage) and the other came from a *Prevotella sp.* genome assembled from metagenomic samples [204] (98.88% identity and 95% coverage) (Figure 3-S5B). After these two similar proteins there is a sharp drop in identity with the next closest proteins also coming from *Prevotella* hosts (Figure 3-S5B). Bacteria from the genus *Prevotella* are generally considered to be commensal to humans, so colistin resistance in this genus is not particularly troublesome. However, continued usage of colistin increases the selection pressure and possibility of horizontal gene transfer of this gene to pathogenic strains. Indeed, *mcr*-1 originated from non-pathogenic species within the genus *Moraxella* and has since spread to human and animal pathogens in the *Enterobacteriaceae* family [205-207]. The identification of this PAP2 colistin resistance gene from a functional metagenomic selection allows time for the characterization and surveillance of this gene before it is transmitted to human or animal pathogens.

3.3.5 Captive ape microbiomes and resistomes cluster closely with non-Westernized humans

To contextualize the microbiomes from our cohorts within a diverse group of human lifestyles and environments, we compared our samples to publicly available human and ape gut microbiomes. We included sequencing data that also targeted the V4 region of the 16S rRNA gene from Westernized and non-Westernized humans. Westernized humans include data from humans in the USA [144] (n = 20) and a peri-urban slum of Lima, Peru [163] (n = 20). The non-Westernized humans can be grouped into a rough gradient of Westernization with the human cohort from this study in the Republic of the Congo (n = 81), rural farmers in El Salvador [163] (n = 20) and farmers in Malawi [144] (n = 20) having intermediate Westernization, and Hadza hunter-gatherers [170] (n = 20) and Guahibo Amerindians [144] (n = 20) being the least Westernized. We also included a study of great ape microbiomes that used the V4 region of the 16S rRNA gene that analyzed captive gorillas from zoos in the USA, France, and Switzerland [175] (n = 8). Analysis of the microbial composition of the gut microbiome by PCoA resulted in Westernized and non-Westernized humans clustering by lifestyle despite differences in geographic location. Contrary to our initial hypothesis that non-human primates would cluster with humans living near them, captive apes clustered near the non-Westernized humans that shared a similar lifestyle, and wild apes clustered separately (Figures 3-4A, 3-S6A). Some samples of the published captive gorilla cohort clustered with our captive gorillas while others were intermediate between captive and wild apes. Alpha diversity, as measured by ASV richness and Shannon diversity, demonstrated no apparent pattern of westernization, but USA humans did have the lowest richness and among the lowest Shannon diversity, as has been demonstrated previously (Figure 3-4B, 3-S6C) [142, 183, 208]. Overall, we observed that the patterns of clustering we observed in our initial analysis are robust to further differences in host geography.

We accessed published shotgun metagenomic sequencing data from similar human and gorilla cohorts for resistome and functional pathway analysis. Westernized human sequencing data was obtained from humans in Norman, USA [146] (n = 20) and the peri-urban slum in Lima, Peru [163] (n =14). Non-Westernized humans included the Tunapuco community in the Peruvian Andean highlands [146] (n = 10), rural farmers in El Salvador [163] (n = 9), Hadza hunter-gatherers from Tanzania [147] (n = 20), and humans from this study in the Republic of the Congo (n = 81). We also included a cohort of wild gorillas (n = 19) from the Sangha region of the Republic of the Congo, a sample site with higher human disturbance [140] than our study site that is ~200 km away. By obtaining shotgun metagenomic sequences we were able to compare differences in the gut microbiome functional pathways and resistome.

Functional pathways were robust to differences in geographic location with captive apes continuing to cluster between wild apes and humans (Figures 3-4C, 3-S6C). All human cohorts clustered closely together despite coming from different countries throughout the world. In fact, there was no significant difference in the compositions of the Hadza and Tunapuco cohorts (pairwise adonis, $P = 0.08$) and between the Tunapuco and El Salvador cohorts (pairwise adonis, $P = 0.06$). Wild chimpanzees and wild gorillas also had no significant differences in gut microbiome composition (pairwise adonis, $P = 0.42$). The gorillas from the Sangha region generally clustered with wild ape samples from this study with a few samples clustering separately, possibly representing samples from different seasonal diets not represented in our data [140]. Overall, all samples had decreased Bray-Curtis distances compared to taxonomic distances, most likely due to redundant functional pathways found across microbial taxa (Figure 3-S6C).

Resistome composition generally clustered by degree of Westernization but had less clear separation between Westernized and non-Westernized humans than taxonomic analysis (Figures

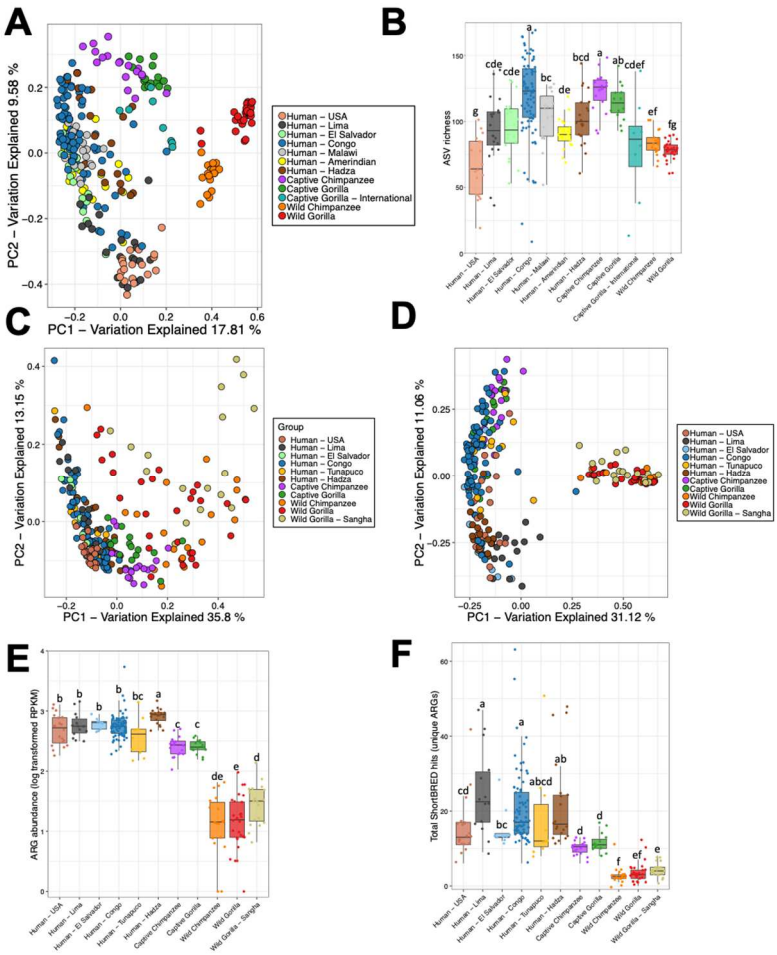


Figure 3-4: Comparison of chimpanzee and gorilla microbiomes to human microbiomes across a wide range of lifestyle and environment. (A) PCoA plot of Bray-Curtis 16S metagenomic profiles. (B) ASV richness of host human, chimpanzee, and gorilla microbiomes (Kruskal-Wallis sum-rank test $P < 2.2e\text{-}16$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 0.021$). (C) PCoA plot using Bray-Curtis distances of the relative abundance of functional pathways from HUMAnN2 results. (D) PCoA plot of Bray-Curtis distances of ARG abundance normalized with RPKM from ShortBRED results. Two wild chimpanzee samples and one wild gorilla sample were removed from this PCoA analysis because the lower rarefaction level used for comparison with public data resulted in no ARGs detected for these samples and no Bray-Curtis distances could be calculated. (E) Sum of all ARG marker abundances per metagenome expressed in log transformed RPKM (Kruskal-Wallis sum-rank test $P < 2.2e\text{-}16$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 0.034$). (F) Resistome richness per metagenome (Kruskal-Wallis sum-rank test $P < 2.2e\text{-}16$; Wilcoxon rank sum test with Benjamini-Hochberg correction $P \leq 0.049$).

3-4D, 3-S6D). Captive chimpanzee and captive gorilla resistomes were not significantly different from each other (pairwise adonis, $P = 0.23$) and the wild chimpanzee, wild gorilla and Sangha

gorilla resistomes were not significantly different from each other (pairwise adonis, wild chimpanzee vs. wild gorilla, $P = 0.63$, wild chimpanzee vs. Sangha gorilla, $P = 0.18$, wild gorilla vs. Sangha gorilla, $P = 0.10$) demonstrating that host species is less influential in resistome composition than lifestyle.

Hadza hunter-gatherers had the highest ARG abundance of all groups analyzed, followed by the other human cohorts, captive apes, and then finally wild apes with the lowest abundance (Figure 3-4E). All the human cohorts from developing countries had high ARG richness with wide variance except for the El Salvador human cohort (Figure 3-4F). Antibiotic resistance may be high in developing countries for a number of reasons, including: a high burden of infectious disease, poor sanitation, and less regulation of antibiotic compounds [209]. Wild apes had the lowest levels of ARGs with captive apes also generally having lower levels of ARG richness and abundance than humans. Lifestyle may be the most influential factor of resistome composition, but host species appears to play a role in ARG alpha-diversity metrics.

In conclusion, contrary to our hypothesis sympatric hosts did not have the most similar gut microbiome composition. Host lifestyle appears to be more influential than host geography or host species at structuring taxonomic and resistome composition. However, functional pathway composition and alpha-diversity metrics of taxonomy and the resistome appear to be influenced by host species and geography. Further, captive ape microbiomes clustered with non-Westernized humans despite living in close proximity to Westernized humans. Our findings indicate that the microbiome of closely related host species may be molded by changes in diet and the degree of antibiotic usage despite the geographic location.

3.4 Materials and Methods

3.4.1 Sample Collection

Wild ape fecal samples were collected in the Goualougo Triangle region of Nouabalé-Ndoki National Park in the Republic of the Congo. Samples were collected in the morning from overnight nesting sites from habituated chimpanzees and gorillas and matched to metadata for the animal that occupied the nest. Human samples were collected from humans living outside of the national park boundaries, including and surrounding the village of Bomassa. Captive ape samples were collected from the Lincoln Park Zoo and the St. Louis Zoo. All sample collection was performed with informed consent for the human participants and in accordance with ethics approval by le Comité d'Ethique de la Recherche en Sciences de la Santé (CERSSA) in the Republic of the Congo. Samples were kept in cold storage on dry ice until shipment to Washington University in St. Louis, where they were stored at -80 °C until processing.

3.4.2 DNA Extraction and Library Preparation

DNA was extracted from 100 mg of each fecal sample using a PowerSoil DNA Isolation Kit (MoBio) with a Mini-beadbeater (BioSpec Products) for the lysis step and stored at -20 °C. Target metagenomic sequencing libraries were created by amplifying the V4 region of 16S rRNA genes using 515F-806R barcoded primers and PCR protocols described previously [210]. Barcoded amplicons were pooled and sequenced on an Illumina MiSeq sequencer with 2 × 250-bp paired-end reads. Shotgun metagenomic sequencing libraries were prepared by diluting extracted DNA to 0.5 ng/µl and using a Nextera DNA Library Prep Kit (Illumina) following the modifications described in Baym et al. to generate ~450 bp segments [122]. DNA fragments were then purified using the Agencourt AMPure XP system (Beckman Coulter) and quantified using

the Quant-iT PicoGreen dsDNA assay (Invitrogen). For each sequencing lane, 10 nM of approximately 80 samples were pooled three independent times. The resulting pools were quantified using the Qubit dsDNA BR Assay and equimolar concentrations were combined. Samples were sequenced on an Illumina NextSeq High-Output platform at the Edison Family Center for Genome Sciences and Systems Biology at Washington University School of Medicine in St. Louis, with a target sequencing depth of 4 million 2 × 150-bp paired-end reads per sample.

3.4.3 16S and shotgun metagenomics bioinformatics analysis

Barcoded 16S Illumina reads were demultiplexed and barcodes were removed using QIIME v1.9 [174]. No significant differences were found in sequencing depth between host cohorts, so samples were not rarefied (Kruskal-Wallis, $P = 0.18$; Figure 3-S1C). Phylogenetic analysis was performed using dada2 v1.8 [211] according to the recommended pipeline. The resulting ASVs were assigned taxonomy using the Greengenes database [212] (version 13_8, 97% clusters). Principal Coordinate Analysis plots were generated using Bray-Curtis distances on sequence counts normalized by relative abundance and significance testing was performed using Adonis. No normalization was performed before alpha-diversity analysis. Significance testing for alpha-diversity was performed using the non-parametric Kruskal-Wallis Rank Sum test and the Pairwise Wilcoxon Rank Sum test using the Benjamini-Hochberg correction for multiple testing. Groups were considered significantly different if the resulting *P*-value was less than 0.05.

Shotgun reads were trimmed for quality and to remove sequence adapters using Trimmomatic v0.38 [124] with the following parameters: *trimmomatic-0.38.jar PE -phred33 ILLUMINACLIP: NexteraPE-PE.fa:2:30:10:1:TRUE LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:60*. Host reads were then removed using DeconSeq v0.4.3 [213, 214] by mapping to the respective host genome (GRCh38 for human; *Pan troglodytes*,

GCA_000001515.5, panTro5 for chimpanzee; *Gorilla gorilla gorilla*, GCA_900006655.1, gorGor5 for gorilla). To normalize for sequencing depth, all samples were rarefied to 4 million paired-end reads. For functional pathway analysis, we used HUMAnN2 v0.9.4 [35] to calculate relative abundance of annotated microbial gene pathways in the MetaCyc database [215]. Since HUMAnN2 does not take paired-reads into account, forward and reverse paired-end reads were combined. The resulting output was normalized to copies per million (CPM) (*humann2_renorm_table*) and samples were joined into a combined table (*humann2_join_tables*).

Downstream statistical analysis was performed in R. Bray-Curtis distances and Adonis significance testing were computed using the vegan package [216]. All of the box and whisker plots represent the 1.5×interquartile ranges in the boxes, medians (50^{th} percentiles) are indicated by the bars within the boxes, the whiskers below and above the box indicate the 10^{th} and 90^{th} percentiles, and the outliers beyond the whiskers are indicated with black circles if all data points are not shown. All statistics for boxplot analyses are based on Kruskal-Wallis tests and if significance is indicated two-tailed *P*-values were calculated using the Wilcoxon rank sum test.

3.4.4 Functional metagenomics and analysis

Due to a limited amount of metagenomic DNA available from individual samples, we pooled samples within host groups to create 16 separate functional libraries (Supplementary Table 3-1). Functional metagenomic libraries were prepared and created as previously described [12]. Briefly, small-insertion (~3-6 kb) expression libraries were created from 5 μg of pooled metagenomic DNA in vector pZE21 and transformed into *E. coli* MegaX DH10B electrocompetent cells. We screened libraries on Mueller-Hinton agar plates containing 12 natural and synthetic antibiotics from 6 different antibiotic classes at concentrations empirically determined to inhibit

the growth of non-transformed MegaX cells. PCR was performed on pooled resistant colonies with vector-specific primers, barcoded, and sequenced.

Sequenced reads were filtered, demultiplexed, and assembled into contigs with PARFuMS v1.1 [167]. We removed all contigs from selections where the number of contigs was more than ten times the number of colonies that grew on the selective agar. Contigs smaller than 500 bp were also removed from further analysis and ORFs were annotated using Resfams v1.2 [217]. We identified 887 contigs which resulted in 332 ORFs that could be classified with high confidence as ARGs related to the screened antibiotic class. Using ShortBRED v0.9.4 [36], we clustered identified ARG protein sequences from functional metagenomic selections and ARG protein sequences from the CARD database [188] with a cluster identity of 90% and used Uniref90 as a reference database for marker creation. We generated 3,092 unique markers for 907 ARG clusters. To measure the abundance of these markers, relative abundance tables normalized to RPKM were created using the *shortbred_quantify.py* script to identify ARG markers in the shotgun metagenomic reads from all our samples.

3.4.5 Microbiome and resistome comparisons to published cohorts

For 16S external analysis, we downloaded public sequencing data of the V4 region of the 16S rRNA gene from subsistence farmers in El Salvador [163], Hadza hunter-gatherers [170], Amerindians [144], and cohorts living in the USA [144], Malawi [144], a peri-urban shanty-town in Lima, Peru [163], and captive gorillas from zoos in the USA, France, and Switzerland [175]. Downloaded data was combined with sequencing reads from this study and all 16S DNA sequences were trimmed to 100 bp to match the length of the shortest downloaded sequences. Samples were rarefied to 9,800 single end reads and the same metagenomics analysis was performed as described above.

For shotgun analysis, we downloaded public shotgun sequencing data from subsistence farmers in El Salvador [163], Hadza hunter-gatherers [147], and cohorts living in the USA [146], Tunapuco, Peru [146], a peri-urban shanty-town in Lima, Peru [163], and wild gorillas from the Sangha region of the Republic of the Congo [140]. After removing all samples from individuals less than 3 years old we randomly selected up to 20 samples from each cohort. We checked for adapters using FastQC v0.11.7 [218]. The sequencing reads appeared to be of good quality and contained no adapters, so we removed host reads using DeconSeq and the human reference genome listed above. The public data contained several samples with less than four million reads, the cutoff used for analysis of Congo humans and wild and captive apes above, so samples were rarefied to three million reads and samples with less than three million reads were removed from analysis. DNA sequencing reads were trimmed to 100 bp to match the length of the shortest downloaded sequences. Some cohorts only had single-end reads so only forward reads were used from paired-end data. Functional analysis was performed using HUMAnN2 as described above. Resistome analysis was performed using ShortBRED with the same parameters as above.

3.4.6 Colistin resistance analysis

The PAP2 gene identified in the functional metagenomic selections was synthesized into the pZE21 vector by Synbio Technologies (Monmouth Junction, NJ). The plasmid was transformed into *E. coli* DH10B and MIC testing was performed using a Colistin ETEST strip (bioMérieux Inc) and confirmed in triplicate using ComASP Colistin Broth Microdilution Susceptibility Testing plates (Liofilchem).

3.4.7 Data availability

16S rRNA and shotgun metagenomic Illumina sequencing data of human, chimpanzee, and gorilla samples is available from the Sequence Read Archive (SRA) under accession PRJNA539933. 16S rRNA data is de-multiplexed and trimmed. Shotgun metagenomic data is de-multiplexed, trimmed, quality-filtered, and host reads have been removed.

3.5 Supplemental Figures

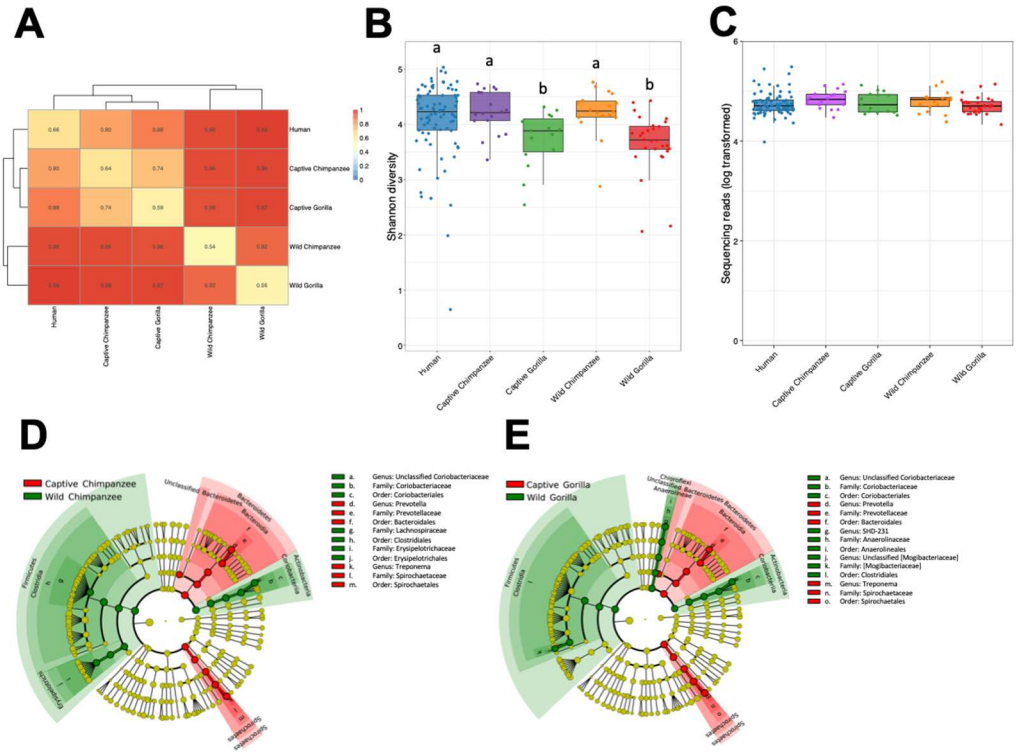


Figure 3-S1: Discriminating taxonomic changes in captive and wild apes. (A) Heatmap of Bray-Curtis distances between host cohorts. (B) Alpha diversity of humans, chimpanzees, and gorillas calculated using Shannon diversity (Kruskal-Wallis P = 8.9e-6; Wilcoxon rank sum test with Benjamini-Hochberg correction P ≤ 0.0076). (C) Boxplot of 16S sequencing depth (Kruskal-Wallis P = 0.18) (D) Cladogram of discriminatory taxa identified in captive and wild chimpanzees (LEfSe LDA log score >4.5, P = 0.05). (E) Cladogram of discriminatory taxa identified in captive and wild gorillas (LEfSe LDA log score >4.5, P = 0.05).

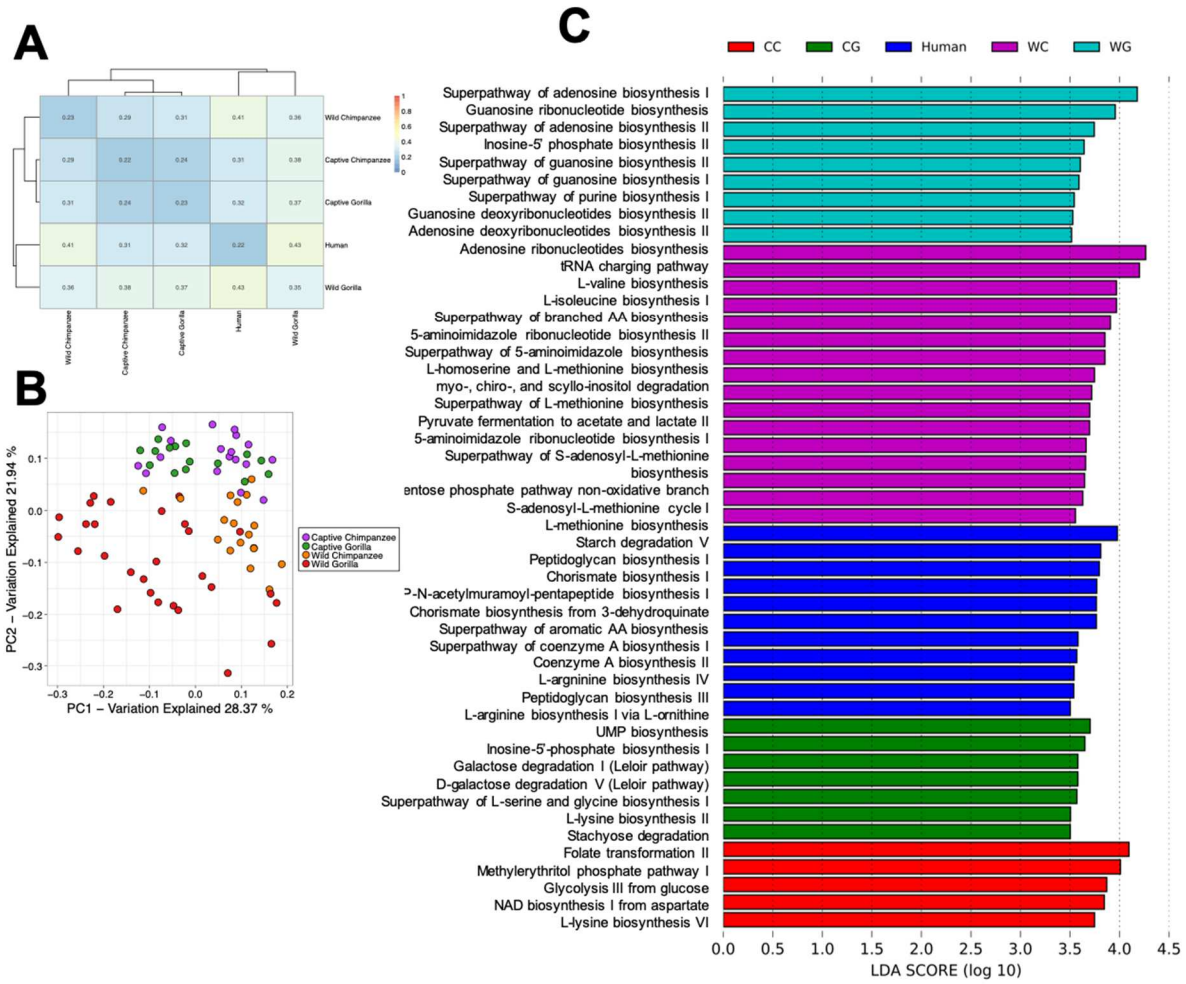


Figure 3-S2: Functional pathway analysis of host microbiota. (A) Heatmap of Bray-Curtis distances of functional pathways. (B) PCoA pot of Bray-Curtis distances of the relative abundance of functional pathways from HUMAnN2 results for chimpanzees and gorillas. (C) Discriminatory functional pathways identified in captive chimpanzees (CC), captive gorillas (CG), humans, wild chimpanzees (WC), and wild gorillas (WG) (LEfSe log score > 3.5, P = 0.05).

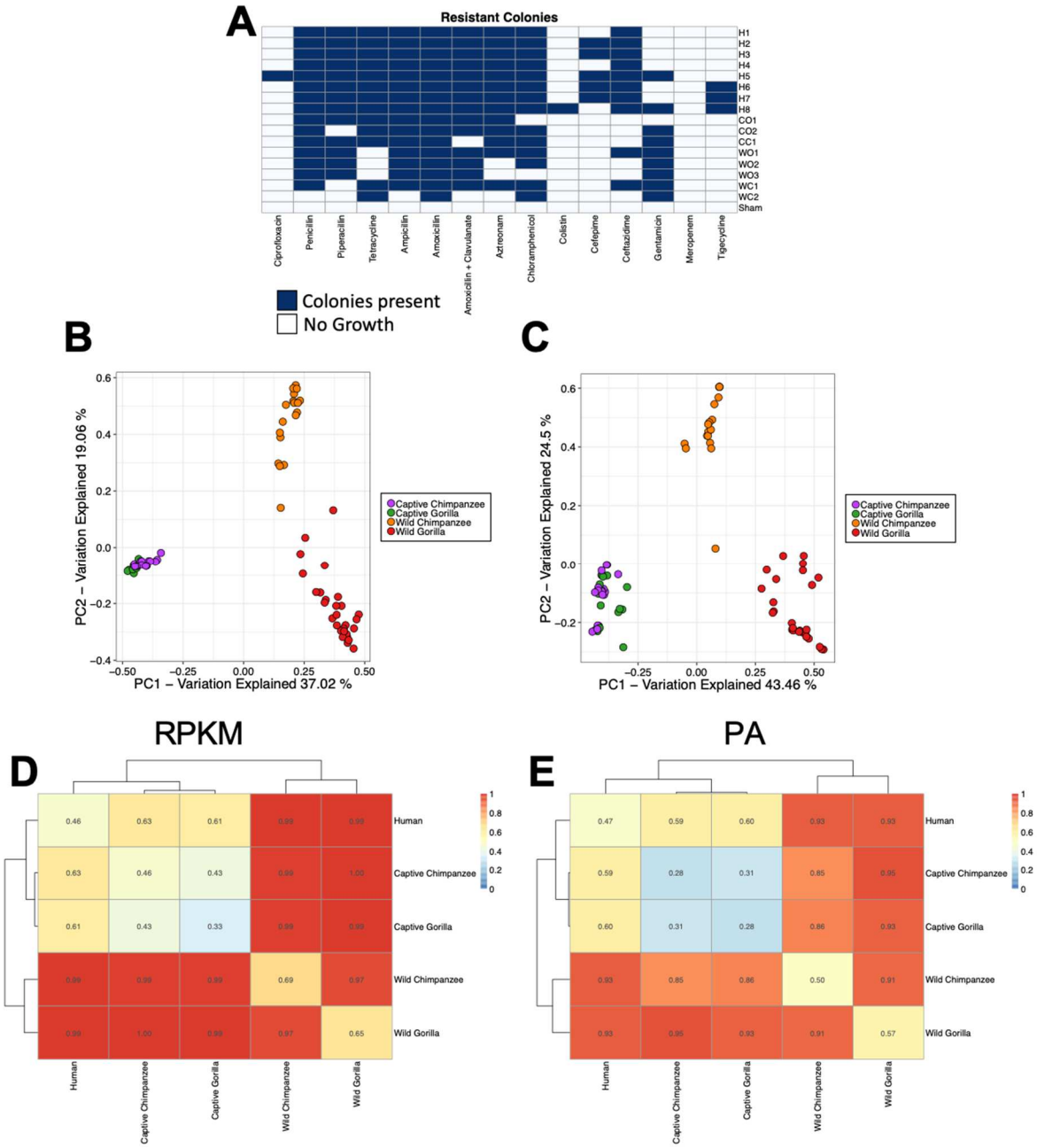


Figure 3-S3: Functional metagenomic resistome analysis of humans chimpanzees and gorillas. (A) Heatmap showing functional metagenome growth results on selective media. Blue squares indicate that colonies were present and white squares indicate no growth was present. Functional libraries are named as described in Supplementary Table 1. (B) PCoA of Bray-Curtis distances of ARG abundance of chimpanzee and gorilla normalized with RPKM. (C) PCoA plot of Bray-Curtis distances of ARG diversity (presence/absence) of chimpanzee and gorilla resistomes. (D) Heatmap of Bray-Curtis distances of ARG abundance using Bray-Curtis distances normalized with RPKM. **e**, Heatmap of Bray-Curtis distances of ARG diversity (presence/absence).

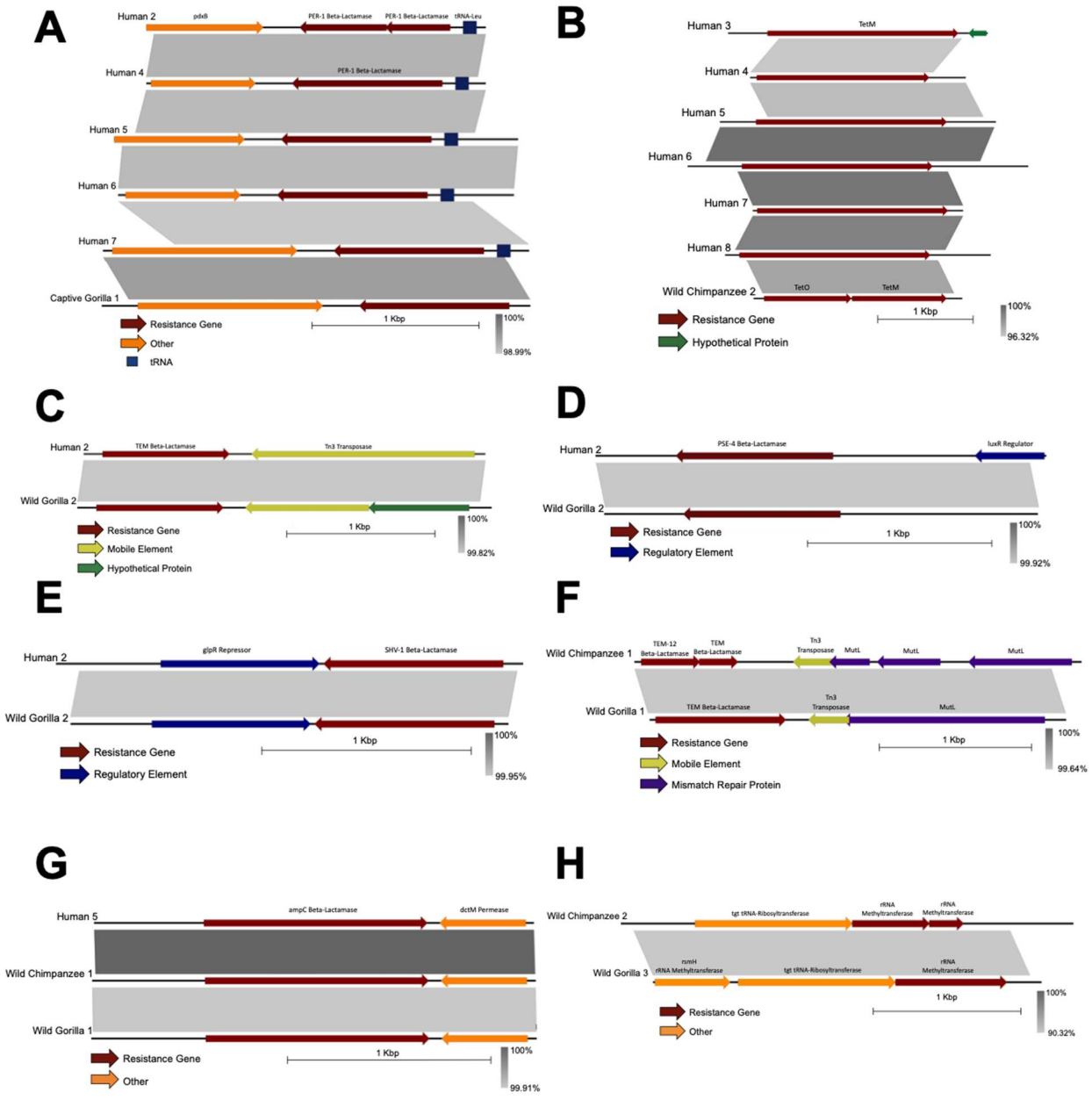


Figure 3-S4: Alignments of ARG containing DNA fragments with high similarity across host cohorts from functional metagenomic selections. (A-H) The library from which the contig originated is indicated.

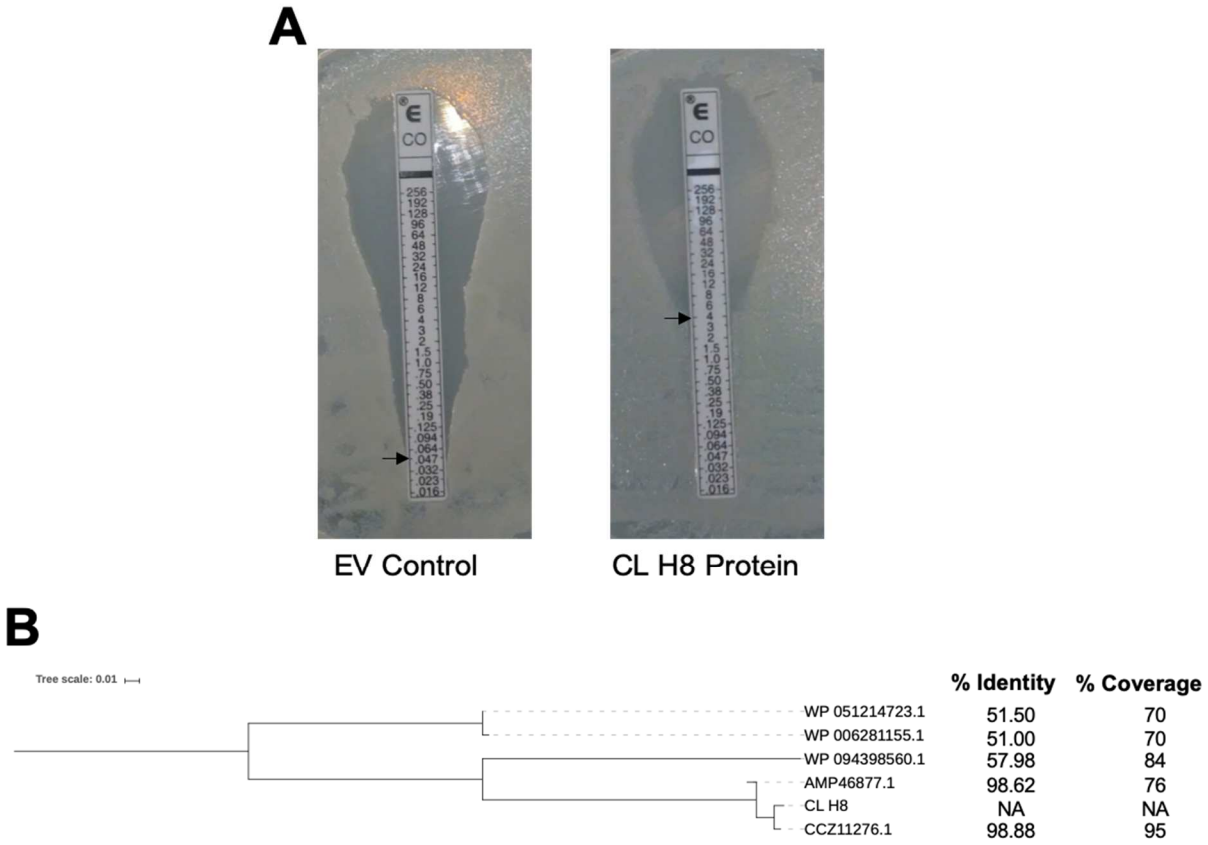


Figure 3-S5: Endogenous expression of colistin ARG is highly resistant to colistin. (A) MIC of colistin against empty vector (EV) control and the novel colistin resistance protein (CL H8). (B) Phylogenetic tree of the colistin resistance protein to the closest BLAST hits from the NCBI non-redundant protein database accessed May 18, 2019. Percent identity and percent coverage are given for each protein in comparison to CL H8.

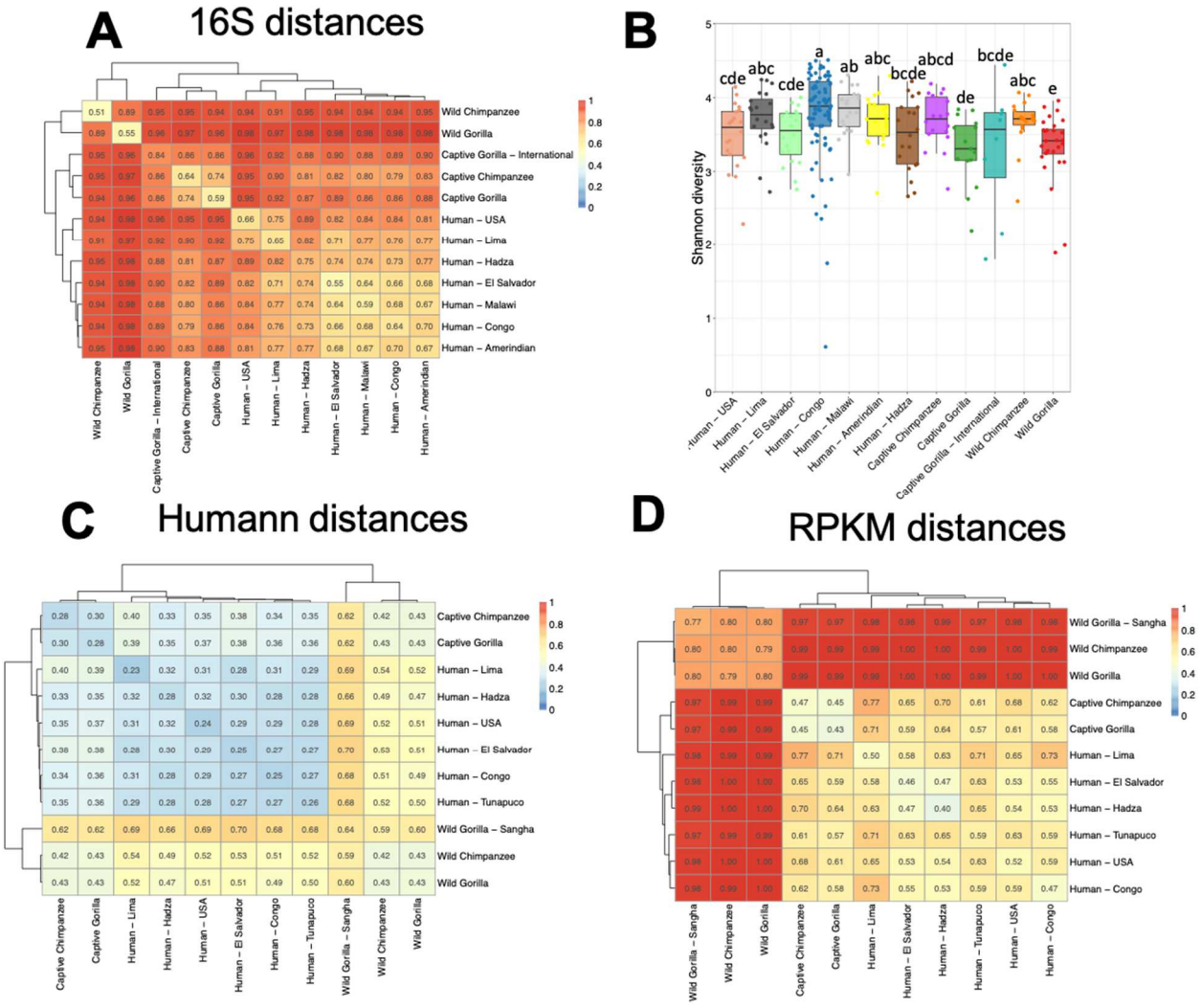


Figure 3-S6: Comparison of ape microbiomes to human microbiomes along a gradient of Westernization. (A) Heatmap of Bray-Curtis distances of 16S microbiome data normalized using relative abundance. (B) Shannon diversity of apes and humans along a gradient of Westernization (Kruskal-Wallis P = 1.5e-6; Wilcoxon rank sum test P $\leq$ 0.049). (C) Heatmap of Bray-Curtis distances of functional pathway abundance normalized using counts per million. (D) Heatmap of Bray-Curtis distances of ARG abundance normalized using RPKM. For this analysis one wild chimpanzee sample had no identified ARGs and was subsequently removed from distance analysis since no Bray-Curtis distance could be calculated.

Table 3-S1: Pooling scheme for functional metagenomic library preparation. Metagenomic DNA was pooled to create 5 μg expression libraries. Extractions with low concentrations required pooling of more samples to achieve a final DNA yield of 5 μg.

Host Species	Samples	Library Name	# of samples included in library	Library Size (GB)
Human	81	H1	11	1.04-2.61
		H2	10	1.26-3.02
		H3	9	2.06-3.13
		H4	10	1.81-5.61
		H5	10	12.37-30.96
		H6	10	5.23-12.68
		H7	9	2.01-5.14
		H8	10	4.18-10.8
Captive Gorilla	15	CG1	7	3.87-5.46
		CG2	8	4.37-6.61
Captive Chimpanzee	17	CC1	17	1.46-4.29
Wild Gorilla	28	WG1	8	2.73-10.73
		WG2	10	0.41-2.51
		WG3	9	2.42-6.23
Wild Chimpanzee	18	WC1	9	3.58-7.05
		WC2	9	1.09-6.53

3.6 Acknowledgements

We thank Dr. Kenneth Cameroon, Wildlife Conservation Society-Congo Program for assistance in the field and shipping of samples from Congo to the USA. Dr. Steve Ross, Maureen Leahy, and the Lester E. Fisher Center great ape staff at Lincoln Park Zoo for their assistance in providing captive ape samples. At the St. Louis Zoo we would like to thank Heidi Helmuth, Jane Merkel, and the great ape care staff. We thank Douglas Berg, Pablo Tsukayama, and Sanket Patel for early discussions regarding study design, and members of the Dantas lab for helpful discussions of the results and manuscript.

This work was supported in part by awards from the National Institutes of Health (NIH) Director's New Innovator Award, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the NIH, and the National Institute of Allergy and Infectious Diseases (NIAID) of the NIH under award numbers DP2DK098089 and R01AI123394 to G.D., the Mallinckrodt Scholar Award of the Edward Mallinckrodt Jr. Foundation to G.D., and the International Center for Energy, Environment, and Sustainability (InCEES) at Washington University in St Louis to C.S. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Chapter 4

Conclusions and Future directions

4.1 Lignocellulose degradation by R. opacus

In Chapter 2, we identified aromatic funneling enzymes and transporters present in *R. opacus* through a combination of RNA-seq and protein identity. Identification of these enzymes allows for the enhancement and possible introduction of these genes into other bacterial strains used for biofuel production. Identification of degradation enzymes across various microbial taxa increases the toolbox available for strain engineering that can then be introduced into wild type *R. opacus* strains in order to optimize biofuel production. This engineering can be further guided by metabolite profiling and metabolic flux analysis that was recently published [219].

Further, through adaptive evolution we found, surprisingly, that none of the mutations found in strains adapted on phenolic compounds were located in genes involved directly in degradation of phenolic compounds. The two genes found in the most strains, superoxide dismutase and cytochrome ubiquinol oxidase subunit I, are involved with redox reactions. Superoxide dismutase quenches radical oxygen species by converting them to either oxygen or hydrogen peroxide molecules, hydrogen peroxide can then be further degraded to oxygen and water [118]. Increased activity in this protein may help counteract the radical oxygen species that are created in aromatic xenobiotic toxicity. However, we found the opposite to be true: the adapted strain containing the mutated superoxide dismutase had increased radical species in the cell

indicating reduced activity in the mutated superoxide dismutase. We hypothesize that the increased amount of superoxide radicals allows for faster oxidation of the phenolic aromatics since degradation enzymes have to generate superoxide radicals to catabolize aromatic compounds [101, 102]. In consort with this, mutations in the cytochrome ubiquinol oxidase subunit I may also increase the concentration of oxygen species within the cell, making them available for aromatic oxidation. Cytochrome ubiquinol oxidase is the last stage of the electron transport chain, and its main function is the final step where electrons are loaded onto oxygen species, reducing them to water [119]. Less efficient activity of this enzyme should thus result in increased oxygen species within the cell but with a cost: some energy is lost from decreased activity of the electron transport chain. However, we did not measure the activity of cytochrome ubiquinol oxidase directly, and it is possible that the mutations introduced into this enzyme enhanced the enzyme activity. Future work can be performed to determine if the introduced mutations increased or decreased enzyme activity.

This study's limitations included the use of model lignin breakdown products instead of real-world lignin and the difficulty of making genetic alterations in *R. opacus*. The aromatic compounds used in this study are only a few of the large moiety of aromatics created during lignin depolymerization. Depolymerized lignin was not used in this study due to limited production capabilities and the tendency of depolymerized lignin mixtures to degrade over time. Although the use of simplified model compounds was not ideal, it provided consistent selective pressure for the adaptation experiments, and the resulting findings can be applied to engineer a strain for lignin depolymerization products. Also, at the time of this project, we did not have the tools necessary to make single nucleotide alterations in the *R. opacus* genome, so we were unable to determine the effect of each individual mutation in isolation. Thus, we don't know which mutation or

combinations of mutations are contributing the most to the increased aromatic tolerance phenotype. Further work is necessary to introduce these mutations into the WT background of *R. opacus* to determine the relative contribution of each of the SNPs that we found in adapted strains. However, characterization of all 255 distinct SNPs is impractical, so initial analysis could be simplified by targeting the SNPs that occurred across more than one adaptation experiment (see Fig. 2-3B).

4.2 Antibiotic resistance in human and ape microbiomes

In Chapter 3, we found that antibiotic treatment and captivity is associated with an increase in antibiotic resistance gene (ARG) richness and abundance. The increased presence of ARGs in antibiotic-treated hosts supports the hypothesis that antibiotic usage selects for resistant bacteria. The abundance of resistance genes increases as resistant bacteria start to dominate the community and as ARGs are disseminated to otherwise susceptible strains through horizontal gene transfer (HGT). We found evidence that some ARGs are shared across host species, even when the hosts are geographically distant. These shared ARGs are found within similar genetic contexts, with some genes being colocalized next to mobile elements, indicating an increased propensity for HGT. These genes may have spread across host microbiomes through transfer from one bacterial strain to another. However, we can't rule out the possibility that these individual ARGs may be found in a bacterial species that is shared across the microbiomes and thus inherited vertically rather than horizontally.

We identified several novel ARGs through our functional metagenomic selections. These genes included resistance to a number of antibiotics, including aminoglycosides and colistin, an antibiotic of last resort [198]. Discovery of novel ARGs provides the opportunity to characterize these genes before they become clinically relevant. The discovery of a new class of tetracycline

resistance genes, termed the tet destructases, from functional metagenomic selections provides a great example of the power of this method. These ARGs were initially discovered in environmental soil samples and have now been characterized [52, 193]. This initial characterization has led to the development of tet destructase inhibitors that rescue tetracycline antibiotics from these resistance genes [220]. Inhibition of tet destructase enzymes is increasingly relevant to human health outcomes because tet destructases have now been found in human pathogens [221, 222]. The discovery and characterization of tet destructases provides an ideal example for how ARG surveillance can lead to clinically relevant outcomes. However, The functionally identified novel resistance genes from this study and several other studies has mostly been ignored with little to no follow up studies on the novel ARGs identified [54, 151, 163, 223]. The colistin resistance gene identified in this study provides a good candidate for further characterization. Future research can be performed to characterize the mechanism of resistance and to determine if this gene can confer colistin resistance in human pathogens such as *Acinetobacter baumannii*. Further work can categorize these novel ARGs by finding all of the ARGs identified in these studies with similar antibiotic targets and mechanisms (e.g. the aminoglycoside methyltransferases or beta-lactamases) and creating phylogenetic trees to visualize the divergence of these genes from known ARGs within the categories. After distinct groupings are identified, representative genes can be assayed to confirm the resistance phenotype and further characterization can be performed on the genes most dissimilar to known resistance genes. This approach will lead to a better understanding of ARG diversity and may provide insight to guide ARG inhibition in the future.

4.3 General Conclusion

In individual isolates equipped with xenobiotic degradation genes, xenobiotics selected for unexpected cellular mechanisms that improved degradation efficiency. These cellular

modifications may be harnessed to create an optimized strain through genetic engineering. In microbial communities with varying levels of xenobiotic tolerance genes, antibiotic selection pressure was associated with an increased abundance of tolerance genes within the community, highlighting the dangers of antibiotic overuse. Overall, knowledge of xenobiotic tolerance in bacteria can be harnessed to benefit industrial processes and protect human health.

References

1. Sahoo, N.K., A. Ramesh, and K. Pakshirajan, *Bacterial Degradation of Aromatic Xenobiotic Compounds: An Overview on Metabolic Pathways and Molecular Approaches*, in *Microorganisms in Environmental Management: Microbes and Environment*, T. Satyanarayana and B.N. Johri, Editors. 2012, Springer Netherlands: Dordrecht. p. 201-220.
2. Top, E.M. and D. Springael, *The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds.* Current Opinion in Biotechnology, 2003. **14**(3): p. 262-269.
3. Anawar, H.M., et al., *Impact of Biochar on Soil Fertility and Behaviour of Xenobiotics in Soil*, in *Xenobiotics in the Soil Environment: Monitoring, Toxicity and Management*, M.Z. Hashmi, V. Kumar, and A. Varma, Editors. 2017, Springer International Publishing: Cham. p. 299-318.
4. Haiser, H.J. and P.J. Turnbaugh, *Developing a metagenomic view of xenobiotic metabolism.* Pharmacol Res, 2013. **69**(1): p. 21-31.
5. Gachanja, A.N. and P.J. Worsfold, *Monitoring of polycyclic aromatic hydrocarbon emissions from biomass combustion in Kenya using liquid chromatography with fluorescence detection.* Science of The Total Environment, 1993. **138**(1): p. 77-89.
6. Hays, M.D., et al., *Speciation of Gas-Phase and Fine Particle Emissions from Burning of Foliar Fuels.* Environmental Science & Technology, 2002. **36**(11): p. 2281-2295.
7. Singh, A. and O.P. Ward, *Biotechnology and Bioremediation — An Overview*, in *Biodegradation and Bioremediation*, A. Singh and O.P. Ward, Editors. 2004, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1-17.
8. Farhadian, M., et al., *In situ bioremediation of monoaromatic pollutants in groundwater: A review.* Bioresource Technology, 2008. **99**(13): p. 5296-5308.
9. Weng, J.K., et al., *Emerging strategies of lignin engineering and degradation for cellulosic biofuel production.* Curr Opin Biotechnol, 2008. **19**(2): p. 166-72.
10. Crofts, T.S., A.J. Gasparrini, and G. Dantas, *Next-generation approaches to understand and combat the antibiotic resistome.* Nat Rev Microbiol, 2017. **15**(7): p. 422-434.
11. Boolchandani, M., A.W. D'Souza, and G. Dantas, *Sequencing-based methods and resources to study antimicrobial resistance.* Nature Reviews Genetics, 2019. **20**(6): p. 356-370.
12. Boolchandani, M., S. Patel, and G. Dantas, *Functional Metagenomics to Study Antibiotic Resistance.* Methods Mol Biol, 2017. **1520**: p. 307-329.
13. Shadnia, H. and J.S. Wright, *Understanding the Toxicity of Phenols: Using Quantitative Structure−Activity Relationship and Enthalpy Changes To Discriminate between Possible Mechanisms.* Chemical Research in Toxicology, 2008. **21**(6): p. 1197-1204.
14. Selassie, C.D., et al., *Phenol toxicity in leukemia cells: a radical process?* Chemico-Biological Interactions, 1998. **113**(3): p. 175-190.
15. Kashmiri, Z.N. and S.A. Mankar, *Free radicals and oxidative stress in bacteria.* International Journal of Current Microbiology and Applied Sciences, 2014. **3**(9): p. 34-40.
16. Cohen, G.M. and M. d'Arcy Doherty, *Free radical mediated cell toxicity by redox cycling chemicals.* The British journal of cancer. Supplement, 1987. **8**: p. 46-52.
17. Porter, T.D. and M.J. Coon, *Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms.* J Biol Chem, 1991. **266**(21): p. 13469-72.

18. Hochstein, P., *Futile redox cycling: implications for oxygen radical toxicity.* Fundam Appl Toxicol, 1983. **3**(4): p. 215-7.
19. Aust, S.D., et al., *Free radicals in toxicology.* Contemporary Issues in Toxicology, 1993. **120**: p. 168-178.
20. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiology and molecular biology reviews : MMBR, 2010. **74**(3): p. 417-433.
21. Walsh, C.T. and T.A. Wencewicz, *Prospects for new antibiotics: a molecule-centered perspective.* The Journal Of Antibiotics, 2013. **67**: p. 7.
22. Aminov, R.I., *A brief history of the antibiotic era: lessons learned and challenges for the future.* Frontiers in microbiology, 2010. **1**: p. 134-134.
23. Lewis, K., *Platforms for antibiotic discovery.* Nat Rev Drug Discov, 2013. **12**(5): p. 371-87.
24. Coates, A., et al., *The future challenges facing the development of new antimicrobial drugs.* Nature Reviews Drug Discovery, 2002. **1**(11): p. 895-910.
25. Babouee Flury, B., et al., *Association of Novel Nonsynonymous Single Nucleotide Polymorphisms in ampD with Cephalosporin Resistance and Phylogenetic Variations in ampC, ampR, ompF, and ompC in Enterobacter cloacae Isolates That Are Highly Resistant to Carbapenems.* Antimicrob Agents Chemother, 2016. **60**(4): p. 2383-90.
26. van der Meer, J.R., et al., *Molecular mechanisms of genetic adaptation to xenobiotic compounds.* Microbiological Reviews, 1992. **56**(4): p. 677.
27. Esteve-Núñez, A., A. Caballero, and J.L. Ramos, *Biological degradation of 2,4,6-trinitrotoluene.* Microbiology and molecular biology reviews : MMBR, 2001. **65**(3): p. 335-352.
28. Sander, P., et al., *Degradation of 1,2,4-trichloro- and 1,2,4,5-tetrachlorobenzene by pseudomonas strains.* Appl Environ Microbiol, 1991. **57**(5): p. 1430-40.
29. Spain, J.C. and P.A. Van Veld, *Adaptation of Natural Microbial Communities to Degradation of Xenobiotic Compounds: Effects of Concentration, Exposure Time, Inoculum, and Chemical Structure.* Applied and Environmental Microbiology, 1983. **45**(2): p. 428.
30. Albrich, W.C., D.L. Monnet, and S. Harbarth, *Antibiotic selection pressure and resistance in Streptococcus pneumoniae and Streptococcus pyogenes.* Emerging infectious diseases, 2004. **10**(3): p. 514-517.
31. Antonopoulos, D.A., et al., *Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation.* Infection and Immunity, 2009. **77**(6): p. 2367.
32. Schuster, S., et al., *Cooperation and cheating in microbial exoenzyme production – Theoretical analysis for biotechnological applications.* Biotechnology Journal, 2010. **5**(7): p. 751-758.
33. Palmer, A.C., E. Angelino, and R. Kishony, *Chemical decay of an antibiotic inverts selection for resistance.* Nat Chem Biol, 2010. **6**(2): p. 105-7.
34. Truong, D.T., et al., *MetaPhlAn2 for enhanced metagenomic taxonomic profiling.* Nature Methods, 2015. **12**: p. 902.
35. Franzosa, E.A., et al., *Species-level functional profiling of metagenomes and metatranscriptomes.* Nature Methods, 2018. **15**(11): p. 962-968.
36. Kaminski, J., et al., *High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED.* PLOS Computational Biology, 2015. **11**(12): p. e1004557.

37. Struijs, J., J. Stoltenkamp, and D. van de Meent, *A spreadsheet-based box model to predict the fate of xenobiotics in a municipal wastewater treatment plant.* Water Research, 1991. **25**(7): p. 891-900.
38. Richards, D.J. and W.K. Shieh, *Biological fate of organic priority pollutants in the aquatic environment.* Water Research, 1986. **20**(9): p. 1077-1090.
39. Peralta-Yahya, P.P., et al., *Microbial engineering for the production of advanced biofuels.* Nature, 2012. **488**(7411): p. 320-8.
40. Ibraheem, O. and B.K. Ndimba, *Molecular adaptation mechanisms employed by ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds.* Int J Biol Sci, 2013. **9**(6): p. 598-612.
41. Pienkos, P.T. and M. Zhang, *Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates.* Cellulose, 2009. **16**(4): p. 743-762.
42. Henson, W.R., et al., *Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus.* Metabolic Engineering, 2018. **49**: p. 69-83.
43. Balan, V., D. Chiaramonti, and S. Kumar, *Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels.* Biofuels, Bioproducts and Biorefining, 2013. **7**(6): p. 732-759.
44. Dunlop, M.J., *Engineering microbes for tolerance to next-generation biofuels.* Biotechnol Biofuels, 2011. **4**: p. 32.
45. Alvarez, H.M., et al., *Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630.* Archives of Microbiology, 1996. **165**(6): p. 377-386.
46. Levy, S.B. and B. Marshall, *Antibacterial resistance worldwide: causes, challenges and responses.* Nat Med, 2004. **10**(12 Suppl): p. S122-9.
47. D'Costa, V.M., et al., *Antibiotic resistance is ancient.* Nature, 2011. **477**(7365): p. 457-61.
48. Bhullar, K., et al., *Antibiotic resistance is prevalent in an isolated cave microbiome.* PLoS One, 2012. **7**(4): p. e34953.
49. Knapp, C.W., et al., *Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940.* Environ Sci Technol, 2010. **44**(2): p. 580-7.
50. Clemente, J.C., et al., *The microbiome of uncontacted Amerindians.* Sci Adv, 2015. **1**(3).
51. Van Boeckel, T.P., et al., *Global trends in antimicrobial use in food animals.* Proc Natl Acad Sci U S A, 2015. **112**(18): p. 5649-54.
52. Park, J., et al., *Plasticity, dynamics, and inhibition of emerging tetracycline resistance enzymes.* Nature chemical biology, 2017. **13**(7): p. 730-736.
53. Bush, K., et al., *Tackling antibiotic resistance.* Nat Rev Microbiol, 2011. **9**(12): p. 894-6.
54. Pehrsson, E.C., et al., *Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs.* Front Microbiol, 2013. **4**: p. 145.
55. Sun, Y. and J. Cheng, *Hydrolysis of lignocellulosic materials for ethanol production: a review.* Bioresource Technology, 2002. **83**(1): p. 1-11.
56. Wheeldon, I., P. Christopher, and H. Blanch, *Integration of heterogeneous and biochemical catalysis for production of fuels and chemicals from biomass.* Current Opinion in Biotechnology, 2017. **45**: p. 127-135.
57. Pandey, M.P. and C.S. Kim, *Lignin Depolymerization and Conversion: A Review of Thermochemical Methods.* Chemical Engineering & Technology, 2011. **34**(1): p. 29-41.
58. Bhat, A.H., Y.K. Dasan, and I. Khan, *Extraction of Lignin from Biomass for Biodiesel Production*, in *Agricultural Biomass Based Potential Materials*, K.R. Hakeem, M. Jawaid, and O. Y. Alothman, Editors. 2015, Springer International Publishing: Cham. p. 155-179.

59. Hahn-Hägerdal, B., et al., *Bio-ethanol – the fuel of tomorrow from the residues of today.* Trends in Biotechnology, 2006. **24**(12): p. 549-556.
60. Jones, J.A., et al., *Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids.* Metabolic Engineering, 2016. **35**: p. 55-63.
61. Lee, J.W., et al., *Homo-succinic acid production by metabolically engineered Mannheimia succiniciproducens.* Metabolic Engineering, 2016. **38**: p. 409-417.
62. Pereira, B., et al., *Efficient utilization of pentoses for bioproduction of the renewable two-carbon compounds ethylene glycol and glycolate.* Metabolic Engineering, 2016. **34**: p. 80-87.
63. Qiao, K., et al., *Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism.* Nature Biotechnology, 2017. **35**: p. 173.
64. Xu, P., et al., *Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals.* Proceedings of the National Academy of Sciences, 2016. **113**(39): p. 10848-10853.
65. Yang, J.E., et al., *One-step fermentative production of aromatic polyesters from glucose by metabolically engineered Escherichia coli strains.* Nature Communications, 2018. **9**(1): p. 79.
66. Zhou, Y.J., et al., *Harnessing Yeast Peroxisomes for Biosynthesis of Fatty-Acid-Derived Biofuels and Chemicals with Relieved Side-Pathway Competition.* Journal of the American Chemical Society, 2016. **138**(47): p. 15368-15377.
67. Zhou, Y.J., et al., *Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories.* Nature Communications, 2016. **7**: p. 11709.
68. Ibraheem, O. and B.K. Ndimba, *Molecular adaptation mechanisms employed by ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds.* International journal of biological sciences, 2013. **9**(6): p. 598-612.
69. Ragauskas, A.J., et al., *Lignin Valorization: Improving Lignin Processing in the Biorefinery.* Science, 2014. **344**(6185): p. 1246843.
70. Valdivia, M., et al., *Biofuels 2020: Biorefineries based on lignocellulosic materials.* Microbial Biotechnology, 2016. **9**(5): p. 585-594.
71. Blazeck, J., et al., *Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production.* Nature Communications, 2014. **5**: p. 3131.
72. Bugg, T.D.H. and R. Rahmanpour, *Enzymatic conversion of lignin into renewable chemicals.* Current Opinion in Chemical Biology, 2015. **29**: p. 10-17.
73. Dunlop, M.J., *Engineering microbes for tolerance to next-generation biofuels.* Biotechnology for Biofuels, 2011. **4**(1): p. 32.
74. Fischer, C.R., D. Klein-Marcuschamer, and G. Stephanopoulos, *Selection and optimization of microbial hosts for biofuels production.* Metabolic Engineering, 2008. **10**(6): p. 295-304.
75. Henske, J.K., et al., *Transcriptomic characterization of Caecomyces churrovis: a novel, non-rhizoid-forming lignocellulolytic anaerobic fungus.* Biotechnology for Biofuels, 2017. **10**(1): p. 305.
76. Jin, Y.-S. and J.H.D. Cate, *Metabolic engineering of yeast for lignocellulosic biofuel production.* Current Opinion in Chemical Biology, 2017. **41**: p. 99-106.

77. Keating, D.H., et al., *Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification.* Frontiers in Microbiology, 2014. **5**(402).
78. Le, R.K., et al., *Utilization of simultaneous saccharification and fermentation residues as feedstock for lipid accumulation in Rhodococcus opacus.* AMB Express, 2017. **7**(1): p. 185.
79. Pham, H.L., et al., *Engineering a riboswitch-based genetic platform for the self-directed evolution of acid-tolerant phenotypes.* Nature Communications, 2017. **8**(1): p. 411.
80. Vardon, D.R., et al., *Adipic acid production from lignin.* Energy & Environmental Science, 2015. **8**(2): p. 617-628.
81. Tsitko, I.V., et al., *Effect of Aromatic Compounds on Cellular Fatty Acid Composition of <em>Rhodococcus opacus</em&gt.* Applied and Environmental Microbiology, 1999. **65**(2): p. 853.
82. Kurosawa, K., et al., *High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production.* Journal of Biotechnology, 2010. **147**(3): p. 212-218.
83. Kurosawa, K., S.J. Wewetzer, and A.J. Sinskey, *Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacusfor lignocellulosic fuel production.* Biotechnology for Biofuels, 2013. **6**(1): p. 134.
84. Kosa, M. and A.J. Ragauskas, *Bioconversion of lignin model compounds with oleaginous Rhodococci.* Applied Microbiology and Biotechnology, 2012. **93**(2): p. 891-900.
85. Xie, S., et al., *Advanced Chemical Design for Efficient Lignin Bioconversion.* ACS Sustainable Chemistry & Engineering, 2017. **5**(3): p. 2215-2223.
86. Yoneda, A., et al., *Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630.* Nucleic Acids Research, 2016. **44**(5): p. 2240-2254.
87. Linger, J.G., et al., *Lignin valorization through integrated biological funneling and chemical catalysis.* Proceedings of the National Academy of Sciences, 2014. **111**(33): p. 12013.
88. Rodriguez, A., et al., *Base-Catalyzed Depolymerization of Solid Lignin-Rich Streams Enables Microbial Conversion.* ACS Sustainable Chemistry & Engineering, 2017. **5**(9): p. 8171-8180.
89. Shuai, L., et al., *Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization.* Science, 2016. **354**(6310): p. 329.
90. Van den Bosch, S., et al., *Integrating lignin valorization and bio-ethanol production: on the role of Ni-Al2O3 catalyst pellets during lignin-first fractionation.* Green Chemistry, 2017. **19**(14): p. 3313-3326.
91. Harwood, C.S. and R.E. Parales, *THE β-KETOADIPATE PATHWAY AND THE BIOLOGY OF SELF-IDENTITY.* Annual Review of Microbiology, 1996. **50**(1): p. 553-590.
92. Kurosawa, K., J. Laser, and A.J. Sinskey, *Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors.* Biotechnology for Biofuels, 2015. **8**(1): p. 76.
93. Conrad, T.M., N.E. Lewis, and B.Ø. Palsson, *Microbial laboratory evolution in the era of genome□scale science.* Molecular Systems Biology, 2011. **7**(1): p. 509.

94. Lenski, R.E. and M. Travisano, *Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations.* Proceedings of the National Academy of Sciences, 1994. **91**(15): p. 6808.
95. Chen, Y., et al., *Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630.* Nucleic Acids Research, 2013. **42**(2): p. 1052-1064.
96. Dionisio, F., et al., *The evolution of a conjugative plasmid and its ability to increase bacterial fitness.* Biol Lett, 2005. **1**(2): p. 250-2.
97. Harrison, E. and M.A. Brockhurst, *Plasmid-mediated horizontal gene transfer is a coevolutionary process.* Trends in Microbiology, 2012. **20**(6): p. 262-267.
98. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**(1): p. 25-29.
99. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature Protocols, 2015. **10**: p. 845.
100. Yates, C.M., et al., *SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features.* Journal of Molecular Biology, 2014. **426**(14): p. 2692-2701.
101. Bugg, T.D.H., *Oxygenases: mechanisms and structural motifs for O2 activation.* Current Opinion in Chemical Biology, 2001. **5**(5): p. 550-555.
102. Gatti, D.L., et al., *The mobile flavin of 4-OH benzoate hydroxylase.* Science, 1994. **266**(5182): p. 110.
103. Patrauchan, M.A., et al., *Proteomic Analysis of Survival of Rhodococcus jostii RHA1 during Carbon Starvation.* Applied and Environmental Microbiology, 2012. **78**(18): p. 6714.
104. Gröning, J.A.D., et al., *Gene redundancy of two-component (chloro)phenol hydroxylases in Rhodococcus opacus 1CP.* FEMS Microbiology Letters, 2014. **361**(1): p. 68-75.
105. Solyanikova, I.P., et al., *Benzoate degradation by Rhodococcus opacus 1CP after dormancy: Characterization of dioxygenases involved in the process.* Journal of Environmental Science and Health, Part B, 2016. **51**(3): p. 182-191.
106. DeLorenzo, D.M., W.R. Henson, and T.S. Moon, *Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630.* ACS Synthetic Biology, 2017. **6**(10): p. 1973-1978.
107. Chen, H.-P., et al., *Vanillin Catabolism in Rhodococcus jostii RHA1.* Applied and Environmental Microbiology, 2012. **78**(2): p. 586.
108. Dardas, A., et al., *The demethylation of guaiacol by a new bacterial cytochrome P-450.* Archives of Biochemistry and Biophysics, 1985. **236**(2): p. 585-592.
109. Priefert, H., J. Rabenhorst, and A. Steinbüchel, *Molecular characterization of genes of Pseudomonas sp. strain HR199 involved in bioconversion of vanillin to protocatechuate.* Journal of Bacteriology, 1997. **179**(8): p. 2595.
110. Lessmeier, L., M. Hoefener, and V.F. Wendisch, *Formaldehyde degradation in Corynebacterium glutamicum involves acetaldehyde dehydrogenase and mycothiol-dependent formaldehyde dehydrogenase.* Microbiology, 2013. **159**(Pt 12): p. 2651-62.
111. Yoshida, N., T. Hayasaki, and H. Takagi, *Gene Expression Analysis of Methylotrophic Oxidoreductases Involved in the Oligotrophic Growth of <I>Rhodococcus*

erythropolis</I> N9T-4. Bioscience, Biotechnology, and Biochemistry, 2011. **advpub**: p. 1012012303-1012012303.
112. Orita, I., et al., *The Ribulose Monophosphate Pathway Substitutes for the Missing Pentose Phosphate Pathway in the Archaeon Thermococcus kodakaraensis</em&gt.* Journal of Bacteriology, 2006. **188**(13): p. 4698.
113. Lee, D.-H., et al., *Cumulative Number of Cell Divisions as a Meaningful Timescale for Adaptive Laboratory Evolution of Escherichia coli.* PLOS ONE, 2011. **6**(10): p. e26172.
114. Chu, N.D., et al., *A Mobile Element in mutS Drives Hypermutation in a Marine Vibrio</em&gt.* mBio, 2017. **8**(1): p. e02045-16.
115. Sandberg, T.E., et al., *Evolution of Escherichia coli to 42 °C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations.* Molecular Biology and Evolution, 2014. **31**(10): p. 2647-2662.
116. Sniegowski, P.D., P.J. Gerrish, and R.E. Lenski, *Evolution of high mutation rates in experimental populations of E. coli.* Nature, 1997. **387**(6634): p. 703-705.
117. Kimchi-Sarfaty, C., et al., *A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity.* Science, 2007. **315**(5811): p. 525.
118. Miller, A.F., *Superoxide dismutases: ancient enzymes and new insights.* FEBS Lett, 2012. **586**(5): p. 585-95.
119. Kishikawa, J.-i., et al., *The cytochrome bcc-aa3-type respiratory chain of Rhodococcus rhodochrous.* Journal of Bioscience and Bioengineering, 2010. **110**(1): p. 42-47.
120. Choi, K.Y., G.J. Zylstra, and E. Kim, *Benzoate Catabolite Repression of the Phthalate Degradation Pathway in Rhodococcus sp. Strain DK17.* Applied and Environmental Microbiology, 2007. **73**(4): p. 1370.
121. Karlson, U., et al., *Two independently regulated cytochromes P-450 in a Rhodococcus rhodochrous strain that degrades 2-ethoxyphenol and 4-methoxybenzoate.* Journal of Bacteriology, 1993. **175**(5): p. 1467.
122. Baym, M., et al., *Inexpensive multiplexed library preparation for megabase-sized genomes.* PLoS One, 2015. **10**(5): p. e0128036.
123. Caruccio, N., *Preparation of Next-Generation Sequencing Libraries Using Nextera™ Technology: Simultaneous DNA Fragmentation and Adaptor Tagging by In Vitro Transposition*, in *High-Throughput Next Generation Sequencing: Methods and Applications*, Y.M. Kwon and S.C. Ricke, Editors. 2011, Humana Press: Totowa, NJ. p. 241-255.
124. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics (Oxford, England), 2014. **30**(15): p. 2114-2120.
125. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. **9**: p. 357.
126. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
127. Walker, B.J., et al., *Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.* PLOS ONE, 2014. **9**(11): p. e112963.
128. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nature Genetics, 2011. **43**: p. 491.
129. Nurk, S., et al., *Assembling single-cell genomes and mini-metagenomes from chimeric MDA products.* J Comput Biol, 2013. **20**(10): p. 714-37.

130. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**(1): p. 421.
131. Holder, J.W., et al., *Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development.* PLOS Genetics, 2011. **7**(9): p. e1002219.
132. Götz, S., et al., *High-throughput functional annotation and data mining with the Blast2GO suite.* Nucleic Acids Research, 2008. **36**(10): p. 3420-3435.
133. Jones, P., et al., *InterProScan 5: genome-scale protein function classification.* Bioinformatics, 2014. **30**(9): p. 1236-1240.
134. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2013. **30**(7): p. 923-930.
135. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 2014. **15**(12): p. 550.
136. Madsen, R.B., et al., *Using design of experiments to optimize derivatization with methyl chloroformate for quantitative analysis of the aqueous phase from hydrothermal liquefaction of biomass.* Analytical and Bioanalytical Chemistry, 2016. **408**(8): p. 2171-2183.
137. Folch, J., M. Lees, and G.H. Sloane Stanley, *A simple method for the isolation and purification of total lipides from animal tissues.* J. Biol. Chem., 1957. **226**: p. 497-509.
138. DeLorenzo, D.M., et al., *Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630.* ACS Synthetic Biology, 2018. **7**(2): p. 727-738.
139. DeLorenzo, D.M., et al., *Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630.* ACS Synth Biol, 2018. **7**(2): p. 727-738.
140. Hicks, A.L., et al., *Gut microbiomes of wild great apes fluctuate seasonally in response to diet.* Nature Communications, 2018. **9**(1): p. 1786.
141. Moeller, A.H., et al., *Cospeciation of gut microbiota with hominids.* Science, 2016. **353**(6297): p. 380.
142. Moeller, A.H., et al., *Rapid changes in the gut microbiome during human evolution.* Proc Natl Acad Sci U S A, 2014. **111**(46): p. 16431-5.
143. Clayton, J.B., et al., *Captivity humanizes the primate microbiome.* Proceedings of the National Academy of Sciences, 2016. **113**(37): p. 10376.
144. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**: p. 222.
145. Clemente, J.C., et al., *The microbiome of uncontacted Amerindians.* Science Advances, 2015. **1**(3): p. e1500183.
146. Obregon-Tito, A.J., et al., *Subsistence strategies in traditional societies distinguish gut microbiomes.* Nature Communications, 2015. **6**: p. 6505.
147. Rampelli, S., et al., *Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota.* Curr Biol, 2015. **25**(13): p. 1682-93.
148. Schnorr, S.L., et al., *Gut microbiome of the Hadza hunter-gatherers.* Nature Communications, 2014. **5**: p. 3654.
149. Gomez, A., et al., *Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns.* Cell Rep, 2016. **14**(9): p. 2142-2153.
150. David, L.A., et al., *Host lifestyle affects human microbiota on daily timescales.* Genome Biology, 2014. **15**(7): p. R89.

151. Tsukayama, P., et al., *Characterization of Wild and Captive Baboon Gut Microbiota and Their Antibiotic Resistomes.* mSystems, 2018. **3**(3).
152. Ley, R.E., et al., *Evolution of mammals and their gut microbes.* Science, 2008. **320**(5883): p. 1647-51.
153. Moeller, A.H., et al., *Sympatric chimpanzees and gorillas harbor convergent gut microbial communities.* Genome Res, 2013. **23**(10): p. 1715-20.
154. Rwego, I.B., et al., *Gastrointestinal bacterial transmission among humans, mountain gorillas, and livestock in Bwindi Impenetrable National Park, Uganda.* Conserv Biol, 2008. **22**(6): p. 1600-7.
155. Song, S.J., et al., *Cohabiting family members share microbiota with one another and with their dogs.* eLife, 2013. **2**: p. e00458-e00458.
156. Dunay, E., et al., *Pathogen Transmission from Humans to Great Apes is a Growing Threat to Primate Conservation.* Ecohealth, 2018. **15**(1): p. 148-162.
157. Walsh, P.D., et al., *Catastrophic ape decline in western equatorial Africa.* Nature, 2003. **422**(6932): p. 611-614.
158. Wittemyer, G., et al., *Accelerated Human Population Growth at Protected Area Edges.* Science, 2008. **321**(5885): p. 123.
159. Kondgen, S., et al., *Pandemic human viruses cause decline of endangered great apes.* Curr Biol, 2008. **18**(4): p. 260-4.
160. Gomez, A., et al., *Temporal variation selects for diet–microbe co-metabolic traits in the gut of Gorilla spp.* The Isme Journal, 2015. **10**: p. 514.
161. Fujimura, K.E., et al., *Role of the gut microbiota in defining human health.* Expert Rev Anti Infect Ther, 2010. **8**(4): p. 435-54.
162. King, T. and C. Chamerlan, *Orphan gorilla management and reintroduction: progress and perspectives.* Gorilla Journal, 2007. **34**: p. 21-25.
163. Pehrsson, E.C., et al., *Interconnected microbiomes and resistomes in low-income human habitats.* Nature, 2016. **533**: p. 212.
164. Porse, A., et al., *Genome Dynamics of Escherichia coli during Antibiotic Treatment: Transfer, Loss, and Persistence of Genetic Elements In situ of the Infant Gut.* Front Cell Infect Microbiol, 2017. **7**: p. 126.
165. Ferreiro, A., et al., *Multiscale Evolutionary Dynamics of Host-Associated Microbiomes.* Cell, 2018. **172**(6): p. 1216-1227.
166. D'Costa, V.M., et al., *Antibiotic resistance is ancient.* Nature, 2011. **477**: p. 457.
167. Forsberg, K.J., et al., *The shared antibiotic resistome of soil bacteria and human pathogens.* Science, 2012. **337**(6098): p. 1107-11.
168. Karami, N., et al., *Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics.* J Antimicrob Chemother, 2007. **60**(5): p. 1142-5.
169. Sommer, M.O.A., G. Dantas, and G.M. Church, *Functional characterization of the antibiotic resistance reservoir in the human microflora.* Science, 2009. **325**(5944): p. 1128-1131.
170. Smits, S.A., et al., *Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania.* Science, 2017. **357**(6353): p. 802.
171. Ellis, R.J., et al., *Comparison of the distal gut microbiota from people and animals in Africa.* PLoS One, 2013. **8**(1): p. e54783.

172. Morgan, D. and C. Sanz, *Naive encounters with chimpanzees in the Goualougo Triangle, Republic of Congo.* International Journal of Primatology, 2003. **24**(2): p. 369-381.
173. Wilson, M.L., et al., *Lethal aggression in Pan is better explained by adaptive strategies than human impacts.* Nature, 2014. **513**: p. 414.
174. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
175. McKenzie, V.J., et al., *The Effects of Captivity on the Mammalian Gut Microbiome.* Integr Comp Biol, 2017. **57**(4): p. 690-704.
176. Basabose, A.K., *Diet composition of chimpanzees inhabiting the montane forest of Kahuzi, Democratic Republic of Congo.* Am J Primatol, 2002. **58**(1): p. 1-21.
177. Calvert, J.J., *Food selection by western gorillas (G.g. gorilla) in relation to food chemistry.* Oecologia, 1985. **65**(2): p. 236-246.
178. Elizabeth Rogers, M., et al., *Gorilla diet in the Lope Reserve, Gabon: : A nutritional analysis.* Oecologia, 1990. **84**(3): p. 326-339.
179. Martinez-Medina, M., et al., *Western diet induces dysbiosis with increased E coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation.* Gut, 2014. **63**(1): p. 116-24.
180. Vangay, P., et al., *US Immigration Westernizes the Human Gut Microbiome.* Cell, 2018. **175**(4): p. 962-972.e10.
181. Less, E.H., et al., *Implementing a low-starch biscuit-free diet in zoo gorillas: the impact on health.* Zoo Biol, 2014. **33**(1): p. 74-80.
182. Morgan, D.B. and C.M. Sanz, *Chimpanzee Feeding Ecology and Comparisons with Sympatric Gorillas in the Goualougo Triangle, Republic of Congo*, in *Feeding Ecology in Apes and Other Primates*, G. Hohmann, M.M. Robbins, and C. Boesch, Editors. 2006, Cambridge University Press: Cambridge. p. 97-122.
183. Nishida, A.H. and H. Ochman, *A great-ape view of the gut microbiome.* Nature Reviews Genetics, 2019. **20**(4): p. 195-206.
184. Segata, N., et al., *Metagenomic biomarker discovery and explanation.* Genome Biol, 2011. **12**(6): p. R60.
185. Han, C., et al., *Complete genome sequence of Treponema succinifaciens type strain (6091).* Standards in genomic sciences, 2011. **4**(3): p. 361-370.
186. Soverini, M., et al., *Variations in the Post-weaning Human Gut Metagenome Profile As Result of Bifidobacterium Acquisition in the Western Microbiome.* Front Microbiol, 2016. **7**: p. 1058.
187. Fernandez, C.I. and A.S. Wiley, *Rethinking the starch digestion hypothesis for AMY1 copy number variation in humans.* Am J Phys Anthropol, 2017. **163**(4): p. 645-657.
188. Jia, B., et al., *CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database.* Nucleic Acids Res, 2017. **45**(D1): p. D566-d573.
189. Ramirez, M.S. and M.E. Tolmasky, *Aminoglycoside modifying enzymes.* Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy, 2010. **13**(6): p. 151-171.
190. Wachino, J. and Y. Arakawa, *Exogenously acquired 16S rRNA methyltransferases found in aminoglycoside-resistant pathogenic Gram-negative bacteria: an update.* Drug Resist Updat, 2012. **15**(3): p. 133-48.

191. Galimand, M., P. Courvalin, and T. Lambert, *Plasmid-mediated high-level resistance to aminoglycosides in Enterobacteriaceae due to 16S rRNA methylation.* Antimicrob Agents Chemother, 2003. **47**(8): p. 2565-71.
192. Fair, R.J. and Y. Tor, *Antibiotics and bacterial resistance in the 21st century.* Perspectives in medicinal chemistry, 2014. **6**: p. 25-64.
193. Forsberg, K.J., et al., *The Tetracycline Destructases: A Novel Family of Tetracycline-Inactivating Enzymes.* Chem Biol, 2015. **22**(7): p. 888-97.
194. Jørgensen, T.S., et al., *Current strategies for mobilome research.* Frontiers in microbiology, 2015. **5**: p. 750-750.
195. Sorensen, S.J., et al., *Studying plasmid horizontal transfer in situ: a critical review.* Nat Rev Microbiol, 2005. **3**(9): p. 700-10.
196. Liu, W., et al., *Origin of the human malaria parasite Plasmodium falciparum in gorillas.* Nature, 2010. **467**(7314): p. 420-425.
197. Liu, W., et al., *African origin of the malaria parasite Plasmodium vivax.* Nature Communications, 2014. **5**: p. 3346.
198. Liu, Y.Y., et al., *Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study.* Lancet Infect Dis, 2016. **16**(2): p. 161-8.
199. Schuchat, A., et al., *Active bacterial core surveillance of the emerging infections program network.* Emerging infectious diseases, 2001. **7**(1): p. 92-99.
200. Roca, I., et al., *The global threat of antimicrobial resistance: science for intervention.* New Microbes and New Infections, 2015. **6**: p. 22-29.
201. Morgan, D., et al., *African apes coexisting with logging: Comparing chimpanzee (Pan troglodytes troglodytes) and gorilla (Gorilla gorilla gorilla) resource needs and responses to forestry activities.* Biological Conservation, 2018. **218**: p. 277-286.
202. Junker, J., et al., *Recent decline in suitable environmental conditions for African great apes.* Diversity and Distributions, 2012. **18**(11): p. 1077-1091.
203. Biswas, S., et al., *Colistin: an update on the antibiotic of the 21st century.* Expert Review of Anti-infective Therapy, 2012. **10**(8): p. 917-934.
204. Nielsen, H.B., et al., *Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes.* Nat Biotechnol, 2014. **32**(8): p. 822-8.
205. AbuOun, M., et al., *mcr-1 and mcr-2 variant genes identified in Moraxella species isolated from pigs in Great Britain from 2014 to 2015.* J Antimicrob Chemother, 2017. **72**(10): p. 2745-2749.
206. Poirel, L., et al., *MCR-2-mediated plasmid-borne polymyxin resistance most likely originates from Moraxella pluranimalium.* J Antimicrob Chemother, 2017. **72**(10): p. 2947-2949.
207. Wei, W., et al., *Defining ICR-Mo, an intrinsic colistin resistance determinant from Moraxella osloensis.* PLOS Genetics, 2018. **14**(5): p. e1007389.
208. Martinez, I., et al., *The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes.* Cell Rep, 2015. **11**(4): p. 527-38.
209. Kunin, C.M., et al., *Social, behavioral, and practical factors affecting antibiotic use worldwide: report of Task Force 4.* Rev Infect Dis, 1987. **9 Suppl 3**: p. S270-85.
210. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.* Isme j, 2012. **6**(8): p. 1621-4.

211. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nature Methods, 2016. **13**: p. 581.
212. DeSantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.* Applied and Environmental Microbiology, 2006. **72**(7): p. 5069-5072.
213. Schmieder, R. and R. Edwards, *Fast identification and removal of sequence contamination from genomic and metagenomic datasets.* PLoS One, 2011. **6**(3): p. e17288.
214. Schmieder, R. and R. Edwards, *Quality control and preprocessing of metagenomic datasets.* Bioinformatics, 2011. **27**(6): p. 863-4.
215. Caspi, R., et al., *The MetaCyc database of metabolic pathways and enzymes.* Nucleic Acids Res, 2018. **46**(D1): p. D633-d639.
216. Oksanen, J., et al., *Vegan: community ecology package*. 2019.
217. Gibson, M.K., K.J. Forsberg, and G. Dantas, *Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology.* Isme j, 2015. **9**(1): p. 207-16.
218. Andrews, S., *FastQC: a quality control tool for high throughput sequence data.* 2010.
219. Roell, G.W., et al., *A concerted systems biology analysis of phenol metabolism in Rhodococcus opacus PD630.* Metabolic Engineering, 2019. **55**: p. 120-130.
220. Markley, J.L., et al., *Semisynthetic Analogues of Anhydrotetracycline as Inhibitors of Tetracycline Destructase Enzymes.* ACS Infectious Diseases, 2019. **5**(4): p. 618-633.
221. He, T., et al., *Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans.* Nature Microbiology, 2019.
222. Sun, J., et al., *Plasmid-encoded tet(X) genes that confer high-level tigecycline resistance in Escherichia coli.* Nature Microbiology, 2019.
223. Gibson, M.K., et al., *Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome.* Nature Microbiology, 2016. **1**: p. 16024.